NMR Studies of the Extracellular Domains of a Human G-Protein Coupled Receptor Grafted onto a Stable β-Barrel Scaffold

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Reto Walser

aus

Grüsch

Promotionskomitee

Prof. Dr. Oliver Zerbe (Vorsitz und Leitung der Dissertation)

Prof. Dr. John Robinson

Zürich, 2010

Table of Contents

Chapter II: Properties of the N-terminal domains from Y-receptors probed by NMR spectroscopy

Chapter III: Grafting of extracellular loops of Y-receptors onto a soluble β-barrel scaffold

Chapter IV: Grafting of loops of Y-receptors onto a membrane-embedded β-barrel scaffold

Summary

The communication of cells with their environment is a vital necessity for all organisms. Often the compounds, which are used to transmit a signal through the extracellular space, are not actually entering their target cells. Rather they are binding to their surface, from where the primary signal is relayed to the cell's interior through specialized receptor molecules located in its membraneous envelope. The signal is then transmitted within the cell through a different compound, which is called a secondary messenger. The largest family of cell surface receptors are the so-called G-protein coupled receptors (GPCRs), which relay the signal to the cell's interior through GTP-binding proteins (G-proteins). The signals to which these receptors respond are of vast diversity ranging from light to small organic molecules to large organic molecules, such as peptide hormones. Their broad ligand range hints to their paramount roles in the regulation of many biological functions ranging from vision to the response to growth hormones. This importance is also reflected in the GPCR's pharmacological importance. Roughly 50% of all currently marketed drugs are targeting this receptor subclass.

The advent of structural biology allowed to understand biological processes on an atomic level. Membrane embedded proteins such as the GPCRs are notoriously difficult to characterize structurally and in spite of their tremendous biological importance only five unique crystal structures of GPCRs are available. Nuclear magnetic resonance (NMR) spectroscopy as the second major technique capable of delivering macromolecular structures at atomic resolution hasn't been successfully applied to the structure determination of GPCRs. Alternative routes to structural and functional information on this receptor class is therefore still highly desirable. In my thesis I present the application of two such alternative strategies to a subfamily of the GPCRs called the Y-receptors. These bind to a set of three neuropeptide hormones, which regulate a host of processes from anxiolysis to feeding behavior and memory retention.

We have tried to isolate the structural entities of the receptor, which are responsible for ligand binding and study them outside of the context of their receptor surroundings. While this approach is obviously only able to deliver limited structural and functional information, it is in our view a valid approach, because many proteins – and membrane proteins in particular – are thought of consisting of small, autonomously folding domains. Therefore they can be expected to retain their structural features even when detached from their natural molecular environment. The first approach is called the “segmentation approach” and aims at dissecting the receptor into suitable fragments and studying these fragments individually. This approach has been applied to the N-termini of the four major members of the Y-receptor family. In order to obtain these four 40-50 residue peptides in ^{15}N-labeled form, they had to be expressed recombinantly in bacteria. Because fragments of such limited length tend to be degraded by bacteria, we had to express them as fusions with suitable

partners conferring stability in the bacterial cell. The N-termini of the Y2- and Y5-receptors (N-Y2 and N-Y5) were produced as fusions to the soluble protein ubiquitin and could be liberated from the fusion partner through treatment with a ubiquitin specific protease. The N-termini of the Y1- and Y4-receptors (N-Y1 and N-Y4) showed unspecific fragmentation when expressed as ubiquitin fusions. This fragmentation problem could be solved by using the insoluble ketosteroidisomerase (KSI) as a fusion partner, from which N-Y1 and N-Y4 could be liberated through a tobacco etch virus (TEV) protease cleavage site introduced between the KSI- and the N-Y1/4 sequence. All four N-termini were devoid of any defined structure in aqueous solution. Upon addition of membrane mimicking micelles N-Y1 and N-Y5 rigidified partially and a defined α-helical segment could be observed in N-Y4. For this fragment also a micromolar affinity to the Y4-receptor's natural ligand (the pancreatic polypeptide, PP) could be detected.

The second approach is called the “grafting approach”. Therein we try to transfer the regions of the Y-receptors, which are of interest to us onto a scaffold protein which is more amenable to NMR spectroscopy then the receptors themselves. An ideal scaffold should be as small as possible to simplify NMR spectroscopy, and should present a number of other favorable characteristics, such as high-level overexpression in a easily manageable host such as *E. coli*. Additionally it should present the grafted sequences in a geometrical fashion resembling the one observed in the natural receptor. We have tested two hypothesized scaffold molecules for their tolerance of accepting the extracellular domains of the Y-receptors. The resulting chimera were assayed for the presence of a stable tertiary fold and the capability of interacting with the Y-receptors' natural ligands.

Both selected scaffolds have a structurally conserved eight-stranded β-barrel core and display flexible loops emanating from the β-barrel core. The bacterial lipocalin (Blc) seemed as a promising soluble scaffold based on structural and sequence information and also based on previous protein engineering work. While we could find numerous suitable attachment points for the Y2-receptor's N-terminus, the Blc scaffold proofed to be much less tolerant to directed alterations in its loops. All changes made to the loops resulted in strongly reduced folding capacity.

As a second potential acceptor molecule we evaluated a membrane embedded scaffold, namely the outer membrane protein A (OmpA) from *E. coli.* The rigid transmembrane β-barrel of this protein proofed highly tolerant to large changes in its solvent-exposed loop sequences. We grafted all three extracellular loops of the Y1-receptor onto this β-barrel scaffold in a number or different topological arrangements. The resulting chimera could all be expressed and purified with similar yields as the wildtype OmpA. Even though the refolding efficiency of the chimera was lower than for OmpA, we managed to elaborate refolding procedures giving yields of >50% of folded protein. In two out of four evaluated chimera we were able to clearly see an interaction between the chimeric protein and the Y-receptors' natural ligands. Our results indicate an interaction of these

chimera with the neurohormones, which qualitatively mimics the natural binding profiles of the Y-receptors. Even though the binding is weak when compared with the affinities of the real receptors, interesting conclusions about the qualitative interaction mode between the chimera and the ligand could be obtained: the peptide hormone ligands seem to interact exclusively with their C-terminal part with the chimeric receptors. This is in agreement with biochemical data available on the interaction between the Y-receptors and their ligands. Specificity of the interaction is corroborated by a competition binding assay and interaction studies with neurohormone mutants showing a decreased affinity towards the natural Y-receptors. Only about half of the resonances from the grafted loops are visible in the spectra of our chimeric receptors. No interaction with the neurohormones can be detected on these resonances. Increasing the overall mobility of the grafted loops by flanking them with flexible linkers also didn't significantly increase the number of visible resonances in the spectra of our receptor constructs and slightly decreased the interaction strength between the ligand and the receptor construct.

Overall we think that our easily accessible chimeric receptor system presents interesting possibilities for gaining insight into the interaction of the Y-receptors and possibly other GPCRs with their ligands.

Zusammenfassung

Kommunikation von Zellen mit der Umwelt ist lebenswichtig für alle Organismen. Die Moleküle, welche zur Übertragung eines Signals verwendet werden, dringen meist nicht in ihre Zielzellen ein, sondern sie werden an deren Oberfläche von Rezeptoren gebunden, welche den extrazellulären Stimulus ins Zellinnere weiterleiten. Das primäre Signal wird so in ein Sekundäres übersetzt. Die G-Protein gekoppelten Rezeptoren (GPCRs) sind die grösste Familie von Oberflächenrezeptoren. Der Name dieser Familie kommt daher, dass das Signal via so genannter GTP bindender Proteine (G-Proteine) ins Zellinnere weitergeleitet wird. Mitglieder dieser Rezeptorfamilie sprechen auf eine grosse Palette verschiedener Signale an, welche sich von Licht über kleine organische Moleküle bis hin zu grossen Biomolekülen wie z.B. Peptidhormonen erstreckt. Diese breite Auswahl deutet schon die überragende Bedeutung dieser Rezeptorfamilie für die Regulation vieler biologischer Prozesse an. Diese Wichtigkeit spiegelt sich auch in der pharmakologischen Rolle der GPCRs wider: Ca. 50% aller derzeit verkauften Medikamente zielen auf GPCRs ab.

Das Aufkommen der Strukturbiologie ermöglichte biologische Prozesse auf atomarer Stufe zu verstehen. Die strukturelle Charakterisierung von in der Membran eingebetteten Proteinen wie den GPCRs ist notorisch schwierig. Und so sind trotz ihrer enormen Wichtigkeit bisher nur fünf Kristallstrukturen von GPCRs veröffentlicht. Die Kernspinresonanzspektroskopie (NMR Spektroskopie) als die zweite Technik, welche in der Lage ist, Strukturen von atomarer Auflösung von biologischen Makromolekülen zu liefern, wurde bisher nicht erfolgreich auf GPCRs angewandt. Alternative Wege zu Strukturinformationen von dieser Rezeptorklasse scheinen also erwünscht zu sein. In meiner Doktorarbeit präsentiere ich die Anwendung zweier solcher alternativen Strategien, angewandt auf eine Unterfamilie der GPCRs, nämlich die Y-Rezeptoren. Diese binden eine Gruppe von drei Neuropeptidhormonen, welche eine Anzahl verschiedener Prozesse regulieren – von der Angstlösung über die Steuerung des Essverhaltens bis hin zur Erinnerung.

Wir haben versucht die strukturellen Untereinheiten, welche für die Bindung der Liganden verantwortlich sind, ausserhalb ihres natürlichen Kontexts zu untersuchen. Diese Herangehensweise erlaubt nur beschränkte strukturelle und funktionale Einsichten. Nichtsdestotrotz ist sie unserer Meinung nach sinnvoll, weil viele Proteine – und Membranproteine insbesondere – als Gruppen von sich autonom faltenden Untereinheiten betrachtet werden können. Darum darf man erwarten, dass diese Untereinheiten ihre Struktureigenschaften auch ausserhalb des Kontexts des gesamten Proteins behalten. Der erste Ansatz nennt sich "Segmentationsansatz" und zielt darauf ab, den Rezeptor in geeignete Fragmente zu zerlegen, und diese Fragmente dann individuell zu untersuchen. Diese Technik wurde auf die N-Termini der vier Hauptvertreter der Y-Rezeptoren angewandt. Um diese 40-50 Reste umfassenden Peptide in ^{15}N-markierter Form für NMR Studien

zur Verfügung zu haben, mussten sie rekombinant in Bakterien hergestellt werden. Weil solch kurze Fragmente im Allgemeinen von Bakterien abgebaut werden, mussten wir sie als Fusionskonstrukte mit geeigneten Partnern exprimieren, welche ihnen Stabilität in den Bakterienzellen verleihen sollen. Die N-Termini der Y2- und Y5-Rezeptoren (N-Y2 und N-Y5) wurden als Fusionen mit dem löslichen Protein Ubiquitin hergestellt, und wurden von diesem Fusionspartner mittels einer Ubiquitin-spezifischen Protease abgespalten. Die N-Termini der Y1- und Y4-Rezeptoren (N-Y1 und N-Y4) zeigten unspezifische Fragmentierung, wenn sie als Fusionen mit Ubiquitin exprimiert wurden. Dieses Fragmentationsproblem konnte durch Fusion der N-Termini an das unlösliche Protein Ketosteroidisomerase (KSI) überwunden werden. Die Spaltung von diesem Fusionspartner erfolgte durch eine zwischen KSI und N-Y1/4 eingefügte "Tobacco Etch Virus" (TEV) Proteaseschnittstelle. Keiner der vier N-Termini zeigte eine definierte Struktur in wässriger Lösung. Nach Zugabe eines Membranmimetikums in Form von Mizellen zeigte sich in N-Y1 und N-Y5 eine partielle Rigidisierung und ein klar definiertes α-helicales Segment konnte in N-Y4 beobachtet werden. Ebenfalls für N-Y4 konnte eine mikromolare Affinität gegenüber dem natürlichen Liganden (dem pankreatischen Polypeptid, PP) des Y4-Rezeptors nachgewiesen werden.

Der zweite alternative Ansatz zu struktureller Information über die Y-Rezeptoren nennt sich "Verpflanzungsansatz". Dabei versuchten wir die Domänen der Y-Rezeptoren, welche für die Ligandenbindung verantwortlich sind, auf ein für NMR-Studien leicht zugängliches Grundgerüst zu übertragen. Ein ideales Grundgerüst für sollte möglichst klein sein, um die NMR Spektroskopie zu vereinfachen. Zudem sollte es eine Anzahl anderer Voraussetzungen erfüllen, wie effiziente Expression in *E. coli* und eine günstige Geometrie, welche jener in den natürlichen Rezeptoren ähnlich sein soll. Wir haben zwei solcher Grundgerüste in Bezug auf ihre Fähigkeit, extrazelluläre Sequenzen der Y-Rezeptoren aufzunehmen, getestet. Die resultierenden Chimären wurden hinsichtlich der Ausbildung einer stabilen Tertiärstruktur und ihrer Affinität gegenüber den Liganden der Y-Rezeptoren untersucht. Beide gewählten Grundgerüste haben einen strukturell konservierten achtsträngigen β-Barrel Kern. Das bakterielle Lipocalin (Blc) schien auf Grund von Struktur- und Sequenzinformation und vorhergehender Mutationsstudien ein viel versprechender Kandidat eines löslichen Grundgerüsts zu sein. Während wir zahlreiche Befestigungspunkte für den N-Terminus des Y2-Rezeptors finden konnten, zeigte sich Blc viel weniger tolerant bezüglich gezielter Änderungen in seinen Loops. Alle Änderungen, die darin vorgenommen wurden, haben die Faltung in eine stabile Tertiärstruktur massiv beeinträchtigt.

Als zweiter möglicher Akzeptor haben wir ein in die Membran eingebettetes Grundgerüst evaluiert, nämlich das Outer membrane protein A (OmpA) von *E. coli*. Das rigide transmembranäre β-Barrel Grundgerüst dieses Proteins war sehr tolerant auch gegenüber grossen Veränderungen in seinen

Loops. Wir haben alle drei extrazellulären Loops des Y1-Rezeptors in jeweils verschiedenen Anordnungen auf dieses β-Barrel Gerüst verpflanzt. Die resultierenden Chimären konnten alle gleich effizient exprimiert und aufgereinigt werden wie OmpA selbst. Obwohl die Rückfaltungseffizienz der chimären Rezeptorkonstrukte niedriger war als für OmpA, konnten wir Rückfaltungsprozesse ausarbeiten, welche uns erlaubten mehr als 50% gefaltetes Protein zu erhalten. In zwei der vier untersuchten Rezeptorkonstrukte konnten wir zudem eine deutliche Interaktion mit den natürlichen Liganden der Y-Rezeptoren beobachten. Unsere Resultate zeigen, dass die Rezeptorkonstrukte mit den Neurohormonen in einer Art wechselwirken, welche die natürlichen Bindungsprofile an den Y-Rezeptoren nachahmt. Obwohl die Bindung verglichen mit den Affinitäten gegenüber der echten Rezeptoren schwach ist, konnten interessante Schlüsse über die qualitative Wechselwirkung gezogen werden: Die Peptidhormone interagieren nur mit ihrem C-terminalen Teil mit den Rezeptorkonstrukten. Dies ist in Übereinstimmung mit veröffentlichten biochemischen Daten über die Interaktion zwischen den Y-Rezeptoren und deren Liganden. Die Spezifität der von uns beobachteten Interaktion wird bewiesen durch Kompetitionsexperimente und Interaktionsstudien mit Mutanten der Neurohormone, welche eine verringerte Affinität gegenüber den natürlichen Y-Rezeptoren zeigen. Nur etwa die Hälfte der Resonanzen von Resten der verpflanzten Loops sind in den Spektren der Rezeptorkonstrukte sichtbar, und keine Anzeichen einer Wechselwirkung mit den Neurohormonen kann in diesen Resonanzen ausgemacht werden. Indem wir die verpflanzten Loops mit flexiblen Linkern versehen haben, wurde die Mobilität dieser Loops erhöht. Dies hat allerdings auch nicht zum vermehrten Erscheinen von Resonanzen in den Spektren geführt, sondern eine etwas abgeschwächte Wechselwirkung dieser Konstrukte mit den Neurohormonen wurde beobachtet.

Wir denken, dass unser leicht zugängliches chimäres Rezeptorsystem interessante Möglichkeiten bietet, um unser Verständnis der Interaktion der Y-Rezeptoren und möglicherweise auch anderer GPCRs mit ihren Liganden zu vertiefen.

Chapter I: Introduction

1. Signal transduction in general and introduction to the topic of this thesis

For both unicellular and multicellular organisms signal transduction across the cellular membrane constitutes a vital need. The hydrophobic nature of the interior of the cellular membrane, however, forms an impenetrable/insurmountable barrier to hydrophilic molecules, precluding passive diffusion of such molecules as an effective means of signal transduction. Membrane embedded proteins are necessary to achieve this task, which can either be accomplished by i) binding the signaling molecule on the extracellular side and, through a conformational change of the membrane embedded protein, releasing it on the intracellular side into the cell or ii) by binding the signaling molecule on the extracellular side and relaying the signal to the intracellular side without a transport of the actual signaling molecule across the membrane. In the second mechanism the signal is then propagated inside the cell by so-called "secondary messengers". The membrane embedded protein relaying the signal from the outside to the inside of the cell is called a "receptor". An important class of such signal receptors is called G protein-coupled receptors (GPCRs), because they relay the extracellular signal in the interior of the cell by a so-called G protein (*vide infra*). There are about 1000 genes coding for such receptors in the human genome, representing the major family of human membrane embedded proteins[1].

GPCRs, as many other membrane proteins as well, are very sensitive molecules with respect to their structural and functional integrity, making their study a very challenging task. One sub-class of the GPCRs is believed to interact with their cognate ligands mainly through its surface exposed extracellular parts. Our work is inspired by the idea, that these extracellular parts could be studied separately, i.e. in absence of the hydrophobic transmembrane helices, responsible for most of the problems arising in the work with GPCRs. Several approaches to realizing this idea have been formulated (*vide infra*). The two routes we are following can be called "segmentation approach" and “grafting approach”. They can be briefly summarized as follows: In the segmentation approach, a GPCR is split at suitable positions, mainly at domain boundaries, and the resulting fragments are studied. The grafting approach is based on the idea of transferring the external parts of a GPCR to a scaffold protein, without disturbing its overall global fold. The scaffold approach relies on the assumption that small, preferably soluble proteins exist, which possess a core determining the global fold and some surface exposed loops unimportant to that fold. The resulting chimeric protein could then be - depending on the size of the scaffold - considered as a “minireceptor”, because it

carries all the parts of the receptor hypothesized to be important for ligand binding (hence the term "receptor"), but at the same time shows the favorable characteristics of small, soluble proteins (hence the term "mini"). The resulting minireceptor should be easier to produce and characterize spectroscopically. This idea is exemplified in figure 1.

Error! Referenc

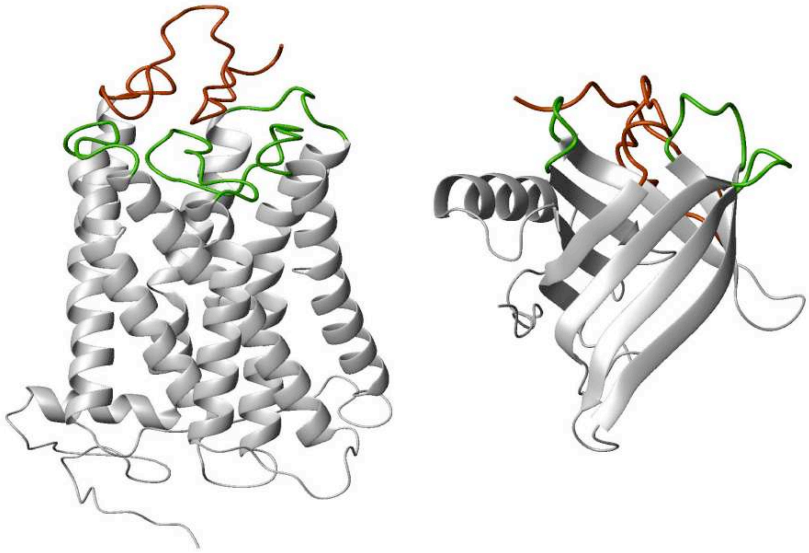


Figure 1: *Exemplification of the grafting approach. The extracellular N-terminus (red) and the loops (green) connecting the transmembrane (TM) helices are colored in a model GPCR on the left. The 7TM scaffold and the incracellular loops are shown in gray. In the selected β-barrel shown on the right might serve as an alternative scaffold for the extracellular domains of a GPCR.*

In the following chapter I would like to set the both basis for understanding the biological concepts most relevant for the study of membrane proteins in general and GPCRs specifically and give an introduction to the experimental techniques we applied.

2. Biological Membranes

2.1 Chemical composition of biological membranes

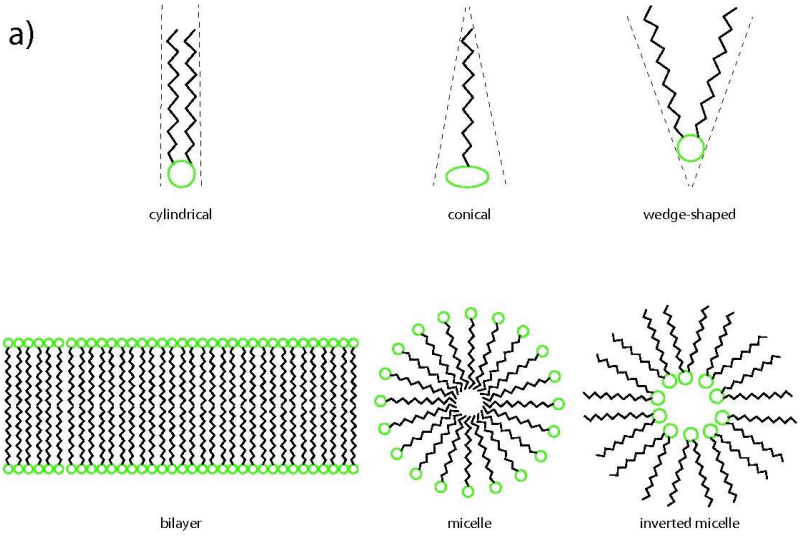
All cells are surrounded by at least one layer of a membrane as a delimiter to the outside world. The two main components of a biological membrane are lipids and proteins. Their ratio varies between different types of membranes, but typically can be considered 1:1. In the inner mitochondrial membrane, which is responsible for maintaining and harvesting the proton-motive force, the protein content is especially high, reaching 75%. In the myelin sheets insulating nerve cells, the protein content is unusually low and only reaches 25%. The functional roles of the two main components of

a biomembrane is considered to be opposite: while lipids form a border against the cell's environment, the proteins are responsible to enable (specific) exchange with the cell's surroundings. Such a communication between a cell's interior and the outside is not only necessary for nutrient uptake, but is essential to allow the coordination between individual cells in a multicellular organism (*vide infra*). Lipids have a characteristic amphiphilicity. They are composed of a hydrophobic hydrocarbon tail and a hydrophilic head. The lipids found in biomembranes can be classified into different families. The glycerolipids have a glycerol backbone which is esterified at two positions with fatty acids of varying length and saturation (*vide infra*). The third hydroxy group is linked to a hydrophilic moiety either through a glycosidic bond to a carbohydrate (glyceroglycolipids) or via a phosphodiester (glycerophospholipids) to an alcohol such as choline, ethanolamine or glycerol. Sphingolipids are derivatives of sphingosine carrying a hydrophobic acyl chain and polar head groups analogous to the glycerolipids. The third major class of biological lipids are the isoprene-based sterols. In higher organisms glycerolipids are mainly found in intracellular organelles. Sphingolipids are most prominent in the plasma membranes of animal cells. Sterols can be found in many different membranes and are absent from bacteria. A structurally and functionally peculiar class of lipids known as lipopolysaccharides are found in the outer membranes of gram negative bacteria (*vide infra*). The acyl chains found in membrane lipids have an even number of carbon atoms ranging from 14 to 26. Predominant are chains with 16 (palmitic acid) and 18 (stearic acid) C-atoms. In case of unsaturated fatty acids the double bond nearly always has the *cis* configuration. The degree of unsaturation varies from 0 to 4 double bonds and in glycerolipids unsaturated chains are usually esterified to position 2 of the glycerol. Palmitoleic acid, the mono-unsaturated form of palmitic acid, can be found in amounts of up to 10%, while higher unsaturated C16 species are very rare. Oleic acid and linoleic acid, the mono- and di-unsaturated forms of stearic acid, make up 10-20% and ~25% of all C18 fatty acid species, respectively. Each membrane has its characteristic lipid composition. Phosphatidylcholine (PC) is abundant in animal membranes, but absent in bacteria, which have a high content of phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). Chloroplasts contain high amounts of glyceroglycolipids and very little phospholipids. Plasma membranes usually have the highest cholesterol content (up to 25%) and in animals are particularly rich in sphingolipids (around 20%). Reflecting its prokaryotic ancestry, the inner mitochondrial membrane is devoid of cholesterol. The two leaflets of a lipid bilayer are usually not symmetric with respect to their lipid composition. This lipid asymmetry can be found in all membranes, but is most pronounced in the plasma membrane. This lipid asymmetry has to be constantly maintained against the tendency of lipids to equilibrate between two leaflets. The dedication of energy resources to the maintenance of lipid bilayer asymmetry indicates its functional importance, which has been observed for instance in the process of apoptosis[2].

2.2 Biophysical properties of biological membranes

All amphiphilic molecules have the tendency to form aggregate structures in solution. In aqueous environments the lipids associate in such a way as to minimize the energetically unfavorable exposure of their fatty acid chains to the polar water molecules. The aggregates can adopt different shapes, depending on the form of the lipids, which is mainly determined by the ratio of the lateral cross sections of the head group and hydrophobic chains. Cone- and wedge-shaped lipids (and also detergents) form micelles or inverted micelles, respectively (figure 1a). Typical glycerophospholipids can be thought of as cylindrical entities with a circular cross-section of ~9 Å (yielding a surface area of ~65 $Å^2$)[3]. In aqueous solution such cylindrical lipids spontaneously organize into bilayers, which gradually curve to form closed vesicles (figure 1a).

The hydrophobicity profile of a bilayer is a smooth transition from the polar extramembrane environment to the apolar membrane interior[4] (figure 2d). The region between these two extremes is generally referred to as the "interface region"[5]. While in the extramembraneous space polar and charged amino acids are predominant and in the hydrophobic core apolar residues prevail, in the interface region aromatic amino acids are found particularly often[6]. In the ~15 Å interface region a gradual transition from the bulk aqueous phase to an essentially anhydrous, apolar hydrocarbon phase takes place[4,7].



b)

Lipid	J_s /nm^{-1}	Ref.
L-lyso PC	1/5.8	a
O-lyso PC	1/3.8	a
P-lyso PC	1/6.8	a
LPA	1/2	b, c
L-lyso PE	<1/40	a
O-lyso PE	<1/40	a
S-lyso PE	<1/40	a
DOPS	1/14.4	d
DOPC	~ -1/10	e, f
PA	-1/4.6	b, c
DOPE	-1/3	g
Cholesterol	~ -1/2.5	e
DAG	~ -1	f, g

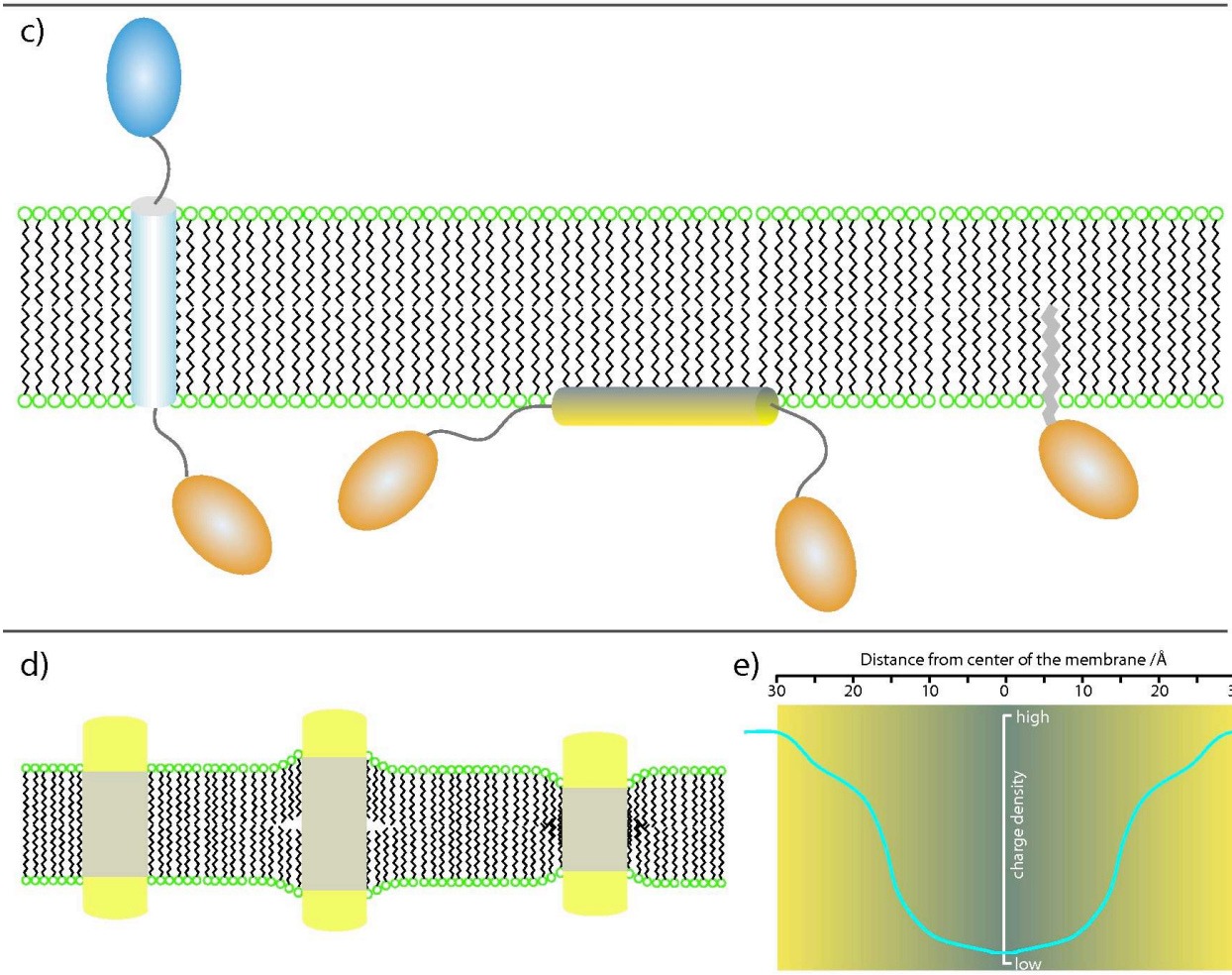


Figure 2: *The biolphysical properties of lipids found in biomembranes. a) Lipids molecules of different shapes aggregate in different forms. b) Lipids with voluminous phatty acid chains (wedge shaped) have negative spontaneous curvature values, whereas such with sterically less demanding chains (cone shaped) have positive spontaneous curvature values. b) The three different modes of membrane protein anchoring: Integral (left), peripheral (middle) and lipid anchored (right). d) Hydrophobic matching: To a certain extent a membrane can adapt its hydrophobi diameter to the specific demands of an embedded protein. d) The hydrophilicity profile through a lipid bilayer is a smooth curve with lowest values in the hydrocarbon core and highest values near the polar headgroups.*

2.3 Proteins embedded in biological membranes

The proteins which are anchored in a membrane can be classified according to different criteria (figure 1c). One of the most commonly used classification is based on the anchoring mechanism, by which the protein is embedded in the membrane. Peripheral membrane proteins are only inserted partially or only interact with the membrane surface, and can usually be detached from the membrane under mild conditions, such as increased ionic strength. Integral membrane proteins on the other hand have one or more domains, which span the full diameter of the membrane, and can usually only be extracted under harsh conditions, e.g. with strong, ionic detergents. Integral membrane proteins can be divided into two structural classes: 1) the ones spanning the membrane with α-helices or 2) the ones inserting into the membrane in the form of a β-barrel. Therefore the α-helix and the β-strand are the only two secondary structural elements, which can be found in membranes. This makes sense, as they ensure the saturation of all hydrogen-bond donors and acceptors in the hydrophobic membrane interior. A minimum of 20 to 30 amino acids in α-helical conformation (1.5 Å rise per residue) are required to pass a biological lipid bilayer, which typically has a thickness of 30-50 Å. Such TM helices are usually highly hydrophobic sequences. This enables their identification of α-helical membrane proteins based on their amino acid sequence. The β-strands of β-barrel membrane proteins, in contrast, are not easy to identify as only every second side chain of a TM β-strand faces the hydrophobic environment of the membrane, whereas the others point into the lumen of the barrel (*vide infra*).

A third class of membrane proteins is anchored in the membrane not through membrane spanning protein sequences, but through posttranslationally attached lipid moieties. The range of such lipid modifications is diverse with the most common being fatty acylation[8] and prenylation[8,9] of cysteines and modifications of C-termini with glycosyl phosphatidyl inositol (GPI)[10,11]. Whereas the former two types of lipid modification is found in the interior of the cell, the latter serves to anchor proteins to the outer leaflet of the plasma membrane.

The diameter of a bilayer is primarily determined by the length and the unsaturation of the fatty acyl chains. It is, however, not a fixed value, but can adjust to environmental conditions. The adaptation of the lipid molecules to match the hydrophobic thickness of an embedded protein has been observed[12,13]. This adaptation is certainly limited to a certain extent. Beyond this limit also a membrane embedded protein can adjust to a given hydrophobic thickness, for instance by tilting its α-helices[14].

3. G protein coupled receptors (GPCRs)

3.1 GPCRs: function

The earliest development of a theory of soluble molecules binding to complementary partners on the cell surface is generally attributed to the german scientist Paul Ehrlich (1854-1915), who in 1897 in his "Seitenkettentheorie" set the basis for today's selection theory in immune response. For this theory he was awarded the nobel price in medicine in 1908. The term "receptor" or "receptive substance" was first mentioned by Langley and his student Dale in the first decade of the 20th century in their work on adreno- and acetylcholine receptor ligands[15,16]. While until the middle of the 20th century the studied receptors were characterized physiologically and pharmacologically (i.e. with smooth or skeletal muscle preparations), in the 1960s and 1970s also biochemical characterization became more and more substantial, resulting in the identification of the intracellular effector machinery of the studied receptors[17-19]. One class, which transmitted their signal through a guanine nucleotide binding protein, was hence called G-protein coupled receptors (GPCRs). For a long time the nicotinic acetylcholine receptor and rhodopsin were the only receptors available in pure forms due to their abundance in electric eel and bovine retina rod cells, respectively. The purification and enrichment of other, less abundant GPCRs only became feasible with the advent of the radioligand binding technique. The purification of the β2-adrenergic receptor or β2-adrenoreceptor (β2-AR)[20] by this method allowed to sequence small peptide stretches of the receptor. From these stretches oligonucleotide probes could be designed that allowed to clone the cDNA of the receptor in 1986[21]. Much to the surprise of everyone the β2-AR showed sequence homology to the visual pigment rhodopsin[21], which was already sequenced in 1982[22,23]. Both receptors were predicted to have a 7 transmembrane topology because of 7 highly hydrophobic stretches present in their sequence. Based on this homology it was quickly suggested that many or all GPCRs might share this structural feature[21] (figure 3a). From then onward many receptor genes were cloned based on homology to the β2-AR or rhodopsin. For these clones often no function was known, and they were hence termed "orphan receptors". Even now, a majority of all cloned GPCRs are still "orphan"[24]. In light of the fact that close to 50% of all currently marketed drugs target GPCRs[25], de-orphanization initiatives are believed to yield attractive new drug targets[26]. After the complete sequencing of the human genome[27] the number of GPCRs in the human genome was estimated to be around 1000[1].

The chemical diversity of signals activating GPCRs is exceptional. It includes molecules as diverse as biogenic amines (dopamine, adrenaline, histamine, acetylcholine, serotonin), peptides (many peptide hormones, some of which are modified [N-formylated, C-amidated], opioids),

glycoproteins, lipids (anandamides, cannabinoids, lysophosphatidic acid), nucleotides (adenosine, cAMP, ADP, ATP) and amino acids (glutamate, GABA). However, GPCRs can not only be activated by chemical signals but also by physical stimuli such as light or mechanical stresses.

All GPCRs act as guanine nucleotide exchange factors (GEFs)[28]. In their activated form GPCRs catalyze the exchange of GDP for GTP on their first downstream effector, the G_α subunit of the so-called heterotrimeric G-proteins. This nucleotide exchange leads to the dissociation of the heterotrimeric G-protein into activated G_α:GTP monomer and a $G_{\beta\gamma}$ heterodimer. Both these molecules can have effector activities. Among other mechanisms (*vide infra*) the duration of the signal is determined by the intrinsic GTP hydrolysis activity of the G_α subunit. GTP hydrolysis of the G_α subunit leads to a termination of the signaling by drastically lowering the G_α activity and by reassociation with the $G_{\beta\gamma}$ subunit[29] (figure 3b).

Heterotrimeric G-proteins are composed of three different subunits named as G_α, G_β and G_γ. In the human genome there are 16 genes that encode G_α-subunits, five genes that encode G_β-subunits and 14 genes that encode G_γ-subunits[30]. The G_α-family is further subdivided into four groups based on sequence homology: $G_{s,\alpha}$, $G_{i/o,\alpha}$, $G_{q/11,\alpha}$ and $G_{12/13,\alpha}$[31]. The first two subfamilies are named according to their effect on adenylyl cyclases. The “s” subscript in $G_{s,\alpha}$ stands for G_α-subunits "stimulating" and the “i” in $G_{i/o,\alpha}$ inhibiting the adenylyl cyclase. Members of the $G_{q/11,\alpha}$ subfamily stimulate phospholipase C_β (PLC-β) and activation of $G_{12/13,\alpha}$ is followed by stimulation of the G-protein Rho[32]. All these "downstream effectors" regulate the activity of further components of the signaling cascade (figure 3b). The large number of effectors involved in a signaling cascade allows the cell to modify (e.g. to amplify or damp) the cellular response at various points. The G_α subunit nomenclature is commonly also used to classify GPCRs: hence, GPCRs are referred to as G_s-, G_i- or G_q-coupled.

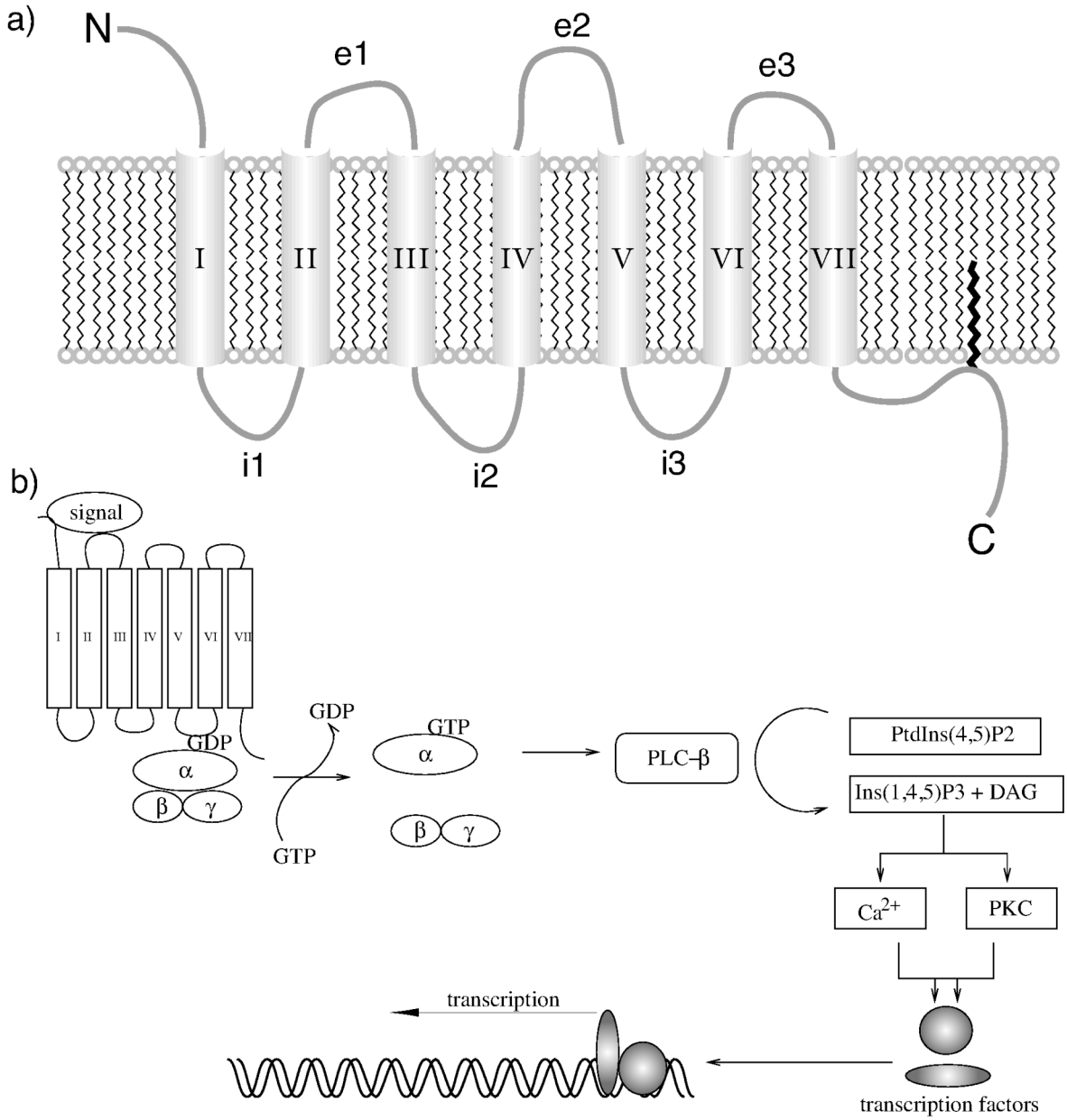


Figure 3: *a) Schematic representation of the structural characteristics of a GPCR. Depicted are the seven transmembrane helices (numbered I to VII) and the extra- and intracellular loops connecting them. Furthermore the extra- and the intracellular N- and C-terminal domains are shown. The C-terminal domain is usually anchored to the lipid membrane through palmitoylation of a cystein residue. b) A possible intracellular signal transduction cascade actviated after the binding of a ligand/signal to a GPCR. Exchange of GDP with GTP on the G_α subunit leads to a dissociation of the heterotrimeric G-protein. The subunits can then activate effectors such as the phospholipase-β (PLC-β) which generates so called "secondary messengers" (e.g. inositol-1,4,5-triphosphate in the case of PLC-β) which in turn lead to cellular responses (e.g. the release of Ca^{2+} from internal stores).*

Within G_α, the C-terminal tail is believed to be the primary site for receptor recognition[33-35], but also six other regions, including the N-terminal domain and several helical regions, within the G_α subunit have been shown to be important for interactions with GPCRs[31,36].

One GPCR might have more than one ligand binding specifically to it, each leading to a different

cellular response. While for instance lipoxin is known to bind selectively to a receptor for formylated peptides[37], leading to an increase in adhesion of cells to laminin, a screening study identified a different class of selective agonists for that receptor, which leads in a different tissue to a transient calcium accumulation in the cytoplasm[38]. This proves the hypothesis that one GPCR can elicit multiple physiological responses.

3.2 GPCRs: structure

Despite the tremendous importance of GPCRs for the (mal)functioning of all higher organisms, fundamental information about their molecular details (e.g. location of ligand binding sites, conformational changes upon activation or the specific interactions with downstream regulatory proteins) is still scarce.

Particularly limiting to biophysical characterization are the naturally low expression levels and high instability of such proteins. For most biophysical characterization techniques, including NMR spectroscopy and X-ray crystallography, large quantities (usually tens of milligrams) of pure protein are required. Because most natural sources have too low expression levels to allow for the extraction of sufficient amounts, recombinant technology is usually applied to increase yields, but also to facilitate the often tedious purification protocols for membrane proteins. However, the heterologous expression systems commonly used (*vide infra*) often lead to heterogenous receptors due to non-homogenous post-translational modifications or misfolding of the receptors in the heterologous hosts.

In the following paragraphs I would like to give a brief overview about the methods used for obtaining membrane proteins in general and specifically GPCRs for biophysical studies. Different units for the quantification of expression levels are in use. The most commonly used are picomoles of the target protein per milligram of total membrane protein (pmol/mg), expressed copies per cell, percent of total membrane protein and percent of total cellular protein. The interconversion between these units for a 45 kDa protein (the typical size of a GPCR) in different host systems is summarized in the following table (all numbers from Schertler, 1992[39]:

	pmol/mg	copies per cell	% of membrane protein	% of total protein
E. coli	10	2000	0.5	0.05
yeast	10	50000	0.25	0.05
insect cells	10	1000000	0.25	0.05
mammalian cells	10	1000000	0.25	0.05

3.2.1 GPCR extraction from naturally (over)producing cells

Rhodopsin is the receptor located in the retina of the eye responsible for the detection of light (it is a photoreceptor). Even though it is a special GPCR in many respects, it is still considered the "prototypical" GPCR, just because it is the best studied thus far (*vide infra*). One peculiarity of rhodopsin is the nature of its stimulus: While most GPCR stimuli are chemical entities (i.e. molecules) it is a physical entity, namely a photon, which leads to the activation of rhodopsin.

The expression of rhodopsin in retinal cells is driven by a strong rod photoreceptor-specific promoter[40]. In human retina, nearly 10^8 opsin molecules are produced per day[41]. In one rod cell there are typically around 10^7 rhodopsin molecules, primarily in the stacked discs but also in the plasma membrane. Activation of 1 to 10^4 photoreceptors can generate a response to low intensity light. The opsins located in the cones (the other major photo-receptive cells of the retina) are responsible for color discrimination. Their overall structure is supposed to be very similar to rhodopsin.

With a chemical homogeneity of 98% the rhodopsin molecules in the rod outer segments (ROS) are uniquely homogenous as compared to other GPCRs[41]. In several studies it was found that the retinal cells' biochemical machinery is also capable of expressing and folding of other GPCRs[42-44]. Rhodopsin carries a characteristic sequence of eight amino acids at its C-terminus. These are necessary for transport of the mature protein to the ROS[45]. This characteristic C-terminal motif has been fused to GPCRs other than rhodopsin in recombinant expression studies of foreign GPCRs in retinal cells (*vide infra*).

3.2.2 Recombinant (over)production

As to now, no universal expression system for the recombinant production of GPCRs has been found or established despite numerous efforts[46,47].

E. coli is by far the most popular host for the expression of recombinant proteins, mostly because of its simplicity of use and the availability of various expression plasmids and strains which have been reported to allow high-level protein production. In *E. coli* there are two strategies for the expression of GPCRs in particular and recombinant proteins in general: (a) expression of functional membrane-inserted receptors or (b) expression of the protein into incorrectly folded, insoluble aggregates, so called "inclusion bodies". The latter approach necessitates the development of a protocol for functional refolding of the receptor after purification under denaturing conditions.

Both approaches depend on the optimization of many parameters. The first strategy was applied to the rat neurotensin receptor NTS1 for example[48]. It has been found that transcription and translation

efficiency were not critical for the overproduction of certain GPCRs, but that receptor insertion into the cytoplasmic membrane seemed to be the rate-limiting step and could be enhanced by expressing the receptor with suitable fusion partners[49-52]. In the example of NTS1 an N-terminal fusion of the *E. coli* periplasmic protein maltose-binding protein (MBP) provided efficient translocation of the MBP by means of its signal peptide, thereby positioning the N-terminus of the fusion construct in the periplasm. At the C-terminus the nature of the affinity tag was found to have a substantial effect on the expression level of the MBP-NTS1 fusion protein. In addition to choosing suitable fusion partners for the protein to be expressed and processed correctly by the translocation machinery "gentle" expression conditions were also required in the case of NTS1 in order not to overload the translocation complex[48]. Such "gentle" conditions can generally be achieved by using low-copy number plasmids with weak promotors and a low growth temperature[53]. The expression levels for such systems are often low, resulting in the production of around 1000 receptor fusion molecules per cell. Subsequent purification requires the careful optimization of buffer composition and detergents used during purification to keep the protein in its native form [53,54]. This includes the design of purification protocols with the least possible number of steps in the shortest possible time to achieve a maximum yield of pure functional receptor.

The approach of expressing a GPCR into insoluble inclusion bodies necessitates the development of a protocol for functional refolding of the receptor[55], a task which sometimes can't be achieved and always requires the screening of a large number of different conditions. This strategy was successfully applied to the human leukotriene B4 receptor[56]. In this study the receptor was purified and refolded in one step by nickel-affinity chromatography under denaturing conditions. After the receptor had bound to the column the denaturing buffer was replaced by a detergent containing buffer, free of denaturant. Refolding efficiency depended, among other factors, on the nature of the detergent, the position of the His-tag, and the column loading[56]. Expression levels for constructs accumulated in inclusion bodies can be very high, reaching >10% of total cell protein.

It should be kept in mind that the overexpression of all eukaryotic proteins in most prokaryotic hosts is limited by the lack of posttranslational modifications and often results in the accumulation of misfolded proteins in inclusion bodies, whether this is desired or not[57].

Additionally the different lipid composition of prokaryotes compared to that of eukaryotes possibly alters GPCR properties[58] and the lack of eukaryote-specific chaperones may prevent proper folding[59,60].

Some prokaryotes, however, do have the capability of performing posttranslational protein modifications. The most notable case is the N-linked glycosylation machinery found in the bacterium *Campylobacter jejuni*. It is encoded by 12 genes organized into a protein glycosylation locus, pgl, and can be functionally transferred to *E. coli* thereby enabling *E. coli* to perform N-

linked glycosylation[61]. In contrast to eukaryotic glycosylation, bacterial N-glycosylation can function on folded proteins and is independent of the translocation process[62]. The principal components of the *C. jejuni* glycosylation system (namely the oligosaccharyltranferase PglB and the lipid-linked oligosaccharide) can be purified and reconstituted to form an efficient *in vitro* protein glycosylation system. Such an *in vitro* system has been used to determine the structure of both the protein and the oligosaccharyl part of the small glycosylated protein AcrA (~100 amino acids) from *C. jejuni*[63].

Yeast provide a eurkaryotic environment which is genetically extremely well characterized and easy to handle even for standard biophysical laboratories. Therefore it combines the ease of genetic manipulation and culturing with the possibility of post-translationally modifying a protein. However, as yeast, representing a "lower" eukaryote, show significantly different glycosylation patterns as compared to "higher" eukaryotes, also in this host misfolding and concomitant malfunctioning of recombinantly expressed proteins of higher eukaryotes has been a problem[64]. Various yeast species have been used for the expression of GPCRs including the baker's yeast *Saccharomyces cerevisiae* and the methylotroph *Pichia pastoris*. In yeast expression systems the foreign gene is either located on an episomal plasmid which replicates independently from the chromosomes, or it lies on a plasmid which integrates into the chromosomes after transfection. Yeasts perform a variety of posttranslational modifications of proteins, they are genetically very well characterized, their growth is usually easy and cheap and large numbers of mutants can be quickly produced and screened. This makes them an interesting production host for recombinant (membrane) protein production. Several GPCRs have been overexpressed to various degrees in *S. cerevisiae*[46,65]. The methylotrophic yeast *P. pastoris* is a particularly interesting expression system because of the high cell densities and the concomitant high expression levels that can be reached with this organism[66]. The expression levels which could commonly be reached for GPCR overexpression in yeast varies widely, but levels of up to >100 pmol/mg have been reported[46,67]. As is the case for bacteria, cultures of simple eukaryotes such as yeast can be grown in media containing up to 100% D_2O. Tuning the the H_2O/D_2O ratio of such cultures allows to achieve almost any desired deuteration level of the target protein[68-70]. Of the lower eukaryotes, especially P. pastoris is an attractive host for protein deuteration due to its good tolerance of D2O and concomitant high cell densities, which thus can be achieved[71].

Even though yeasts are capable of posttranslationally modifying proteins, it should be noted that these modifications can differ from the ones that are found in higher eukaryotes. This is particularly true for N-linked glycosylation. Efforts are undertaken, however, to design yeast strains with a "humanized" glycosylation pattern[72].

The low maintenance requirements and the ability of insect cells to perform posttranslational

modifications has established insect cell cultures as a popular expression system for recombinant eukaryotic proteins more than 25 years ago[73]. Insect cell lines from *Spodoptera fruigperda* (*Sf*) ovarian tissue are the most commonly used insect cell lines[74]. These insect cells are capable of carrying out all posttranslational modifications commonly found in mammalian cells (e.g. phosphorylation, glycosylation, fatty acid acylation). It should be mentioned, however, that the N-glycosylation pattern is more simple than in mammals and usually of the "high-mannose" type[75]. To enhance recombinant protein expression[76,77] and obtain a mammalian-like posttranslational modification pattern insect cell lines from different species, *Trichoplusia ni* and *Estigmena acrea*, were developed[78,79]. Overexpression of the target protein is induced by the infection of the cells with a genetically modified baculovirus, carrying the coding region of the protein to be expressed under a strong viral promotor, usually the polyhedrin promotor[79]. The virus of choice is usually the *Autographa californica* virus, which is a double-stranded DNA virus, surrounded by a lipid membrane[80]. The viral cycle is lytic and causes cell death 4-5 days after infection. Sometimes it can be desirable to express the foreign protein under the control of a weaker early viral promotor. These early promotors, however, show considerably lower expression efficiency.

Baculovirus-infected insect cells have proven efficient for the production of several GPCRs, allowing the overexpression of the β1- and β2-AR and the A_{2A} adenosine receptor, which finally resulted in the solution of their crystal structures (*vide infra*). Some of these could be purified to homogeneity in mg quantities (expression levels of 10-100 pmol/mg) including the 5-HT3 serotonin receptor[81], the human histamine H1 receptor[82], the human thromboxane A2 receptor[83], the human A_{2A} adenosine receptor[84,85], the human cannabinoid CB1 and CB2 receptors[86], and the human β1- and β2-adrenergic receptors[87]. Generally 10-100 pmol of receptor per mg of total membrane protein could be obtained, corresponding to ~10^6 receptors per cell, which is as much as 500 times as high as the expression levels found in the naturally producing mammalian cells[46].

The photoreceptor cells (PRCs) of the *Drosophila melanogaster* eye have been used for the recombinant overexpression of GPCRs[88]. These cells contain a protein synthesis machinery which allows to synthesize rhodopsin in large quantities (*vide supra*). It is possible to genetically engineer *D. melanogaster* to allow for the tissue- and developmental stage specific overexpression of recombinant proteins. The combination of these two factors makes the *Drosophila* PRCs a potentially interesting system for the recombinant production of large quantities of recombinant proteins in general. The system has been used for the overexpression, purification and functional reconstitution of the *D. melanogaster* metabotropic glutamate receptor (DM-GluRA). However, as was also found in this study not all GPCRs could be overexpressed successfully. In the case of the DM-GluRA, however, the system proved to be highly efficient, giving 3-fold higher yields than the commonly used *Sf9* insect cells[88]. The system offers other advantages such as the intrinsically low

protease activity in photoreceptor cells and the high ratio of target disc membranes to other cellular membranes leading to smaller amounts of misfolded receptor due to transport of the receptor to the incorrect location.

Similarly the strategy of expressing GPCRs recombinantly in photoreceptor cells could also be applied to the human serotonin receptor and the endothelial differentiation gene receptor family in *Xenopus laevis* tadpoles[89]. While no data on the glycosylation homogeneity in the *D. melaongaster* expression system is given, Zhang *et al.* report homogenous glycosylation in the *X. laevis* system[89]. This is in contrast to other higher eukaryotic expression systems, for which often heterogenous glycosylation patterns are found. They attribute the lack of un- or mis-glycosylated protein to the fact that the rod photoreceptor cells will only transport properly folded and homogeneously glycosylated proteins to the rod outer segment (ROS). Thereby the rods themselves "purify" the proteins within the ROS[41].

As the main interest is in human GPCRs, expression systems mimicking human cells as closely as possible are generally the most preferred. GPCRs have been recombinantly overexpressed in a number of mammalian cells including human embryonic kidney (HEK) cells, which have been used successfully to overexpress dopamine receptors[90], α- and β-adrenergic receptors[91,92], rhodopsin[93] and yeast pheromone receptor Ste2p[94]. Other popular mammalian cells for recombinant protein production are chinese hamster ovary (CHO) or COS (stemming from monkey kidney) cells. Two fundamentally different approaches have been used to transfer heterologous genetic material into mammalian cells: transient and stable transfection. In transient expression genetic material is injected into the host cells either by transfecting them with a plasmid or by delivery of the genetic material using a recombinant virus. Both techniques use non-selective conditions (i.e. no selection markers) and are therefore usually preferred for short-term protein expression. This is in contrast to stable cell lines where the target gene is co-transfected together with a selectable marker, and application of selective pressure allows to propagate cells having taken up the genetic material over many generations. The transient expression strategy is ideal for functional studies and characterization of a target membrane protein by ligand binding assays or patch-clamp techniques. Usually only a fraction of the cells are transfected and those only survive for a few days. Many membrane proteins have been expressed transiently by transfection of mammalian cells with plasmid DNA and robust overexpression in the 100 pmol/mg range could be observed[95-97]. An alternative method for transiently expressing foreign genes in mammalian cells is provided by recombinant viruses. Compared to the direct transfection with plasmid DNA, the viral systems offer the advantage of high infection efficiency and the presence of strong viral promotors is responsible for high transgene expression levels. The drawbacks of viral vectors lies in their potential bio safety risk and the high cell lethality. The two most commonly used viruses for transfection of mammalian

cells are the *vaccinia* and the *Semliki Forest virus* (SFV). These viruses are capable of infecting a large number of mammalian (and other higher eukaryotic) cell lines[98]. This has allowed the production of recombinant GPCRs in mammalian cells in suspension cultures in many cases. Often the expression levels were up to 100-fold higher than achieved with other systems, lying above 100 pmol/mg[99,100]. Glycosylation of GPCRs produced in SFV-based expression systems is not always homogenous[99] and should therefore be checked carefully.

If a large quantity of a membrane protein is required over a long period of time, it is generally worth the work of establishing a stable cell line. Stable cell lines can be achieved using constructs that integrate into the host genome or replicate and segregate episomally. Integrating constructs are often put under selective pressure to select for clones having incorporated multiple copies of the target gene. For the β2-adrenergic receptor expression levels of 200 pmol/mg could be achieved with this strategy[101].

As deuteration of large proteins is often a necessity to obtain NMR spectra of high quality, it should be kept in mind, that cells of higher eukaryotic species will not grow in media containing more than 30-50% D_2O. The fractional deuteration resulting from this low H_2O/D_2O ratio is often insufficient for NMR studies. This problem can be circumvented by supplementing the growth medium with ^{2}H-labeled amino acids from algal extracts, analogous to the strategies for obtaining ^{15}N- and ^{13}C-labeled proteins in these systems[70,102]. Proteins with a high deuteration levels at all carbon atoms except C_α can thus be produced also in higher eukaryotes.

3.2.3 Cell-free protein expression

Overexpression of (heterologous) protein often faces difficulties ranging from impaired cell-growth due to toxicity of the overexpressed protein to instability of the target protein due to intracellular protein degradation mechanisms. Such issues can be circumvented by so-called cell-free (CF) expression[103], which is an *in vitro* translation system, consisting exclusively of the components required for ribosomal protein synthesis (tRNAs, ribosomes, free amino acids, aminoacyl-tRNA synthetases, nucleoside triphosphates, initiation-, elongation-, and termination factors, ...). Furthermore, CF systems are "open", allowing direct access/manipulation during translation. An additional benefit from an NMR point of view is the absence of cellular metabolic pathways responsible for isotope scrambling[104,105]. In basic translation mixes no membrane compartments are present and the expression of membrane proteins results in the accumulation of insoluble aggregates, which however can often be solubilized by mild detergents[106,107]. The addition of detergents or lipids to the translation mix allows to insert an expressed membrane protein directly from the ribosome into micelles[108,109] or liposomes[110-112].

Extracts of *E. coli* and wheat germs are the most widely used systems. The cell extracts can be

prepared in a normal molecular biology lab[113,114], but are also available commercially[115,116]. Protocols for the preparation of *E. coli*[117-119] and wheat germ[120,121] extracts are well documented. Because of their pharmacological importance and the difficulties in heterologous expression, GPCRs are particularly interesting candidates for CF expression systems[108,122,123].

The protein yields that can be achieved with the CF system can be very high, typically giving mg of protein from 1 ml of the translation reaction. For both soluble[113,117] and membrane proteins[112,124,125] yields in the mg/ml range have been reported. Together with the short reaction times this makes CF expression also an interesting technique for high-throughput proteomics[126,127].

3.2.4 Structural characterization of GPCRs

Circular dichroism is able to report on the secondary structure of proteins. In 1978 studies using this technique provided the first experimental evidence for an extensive α-helical content in rhodopsin, which was consistent with 7-9 transmembrane (TM) helical segments[128]. In 1983 when the primary sequence of rhodopsin was published a hydropathy plot indicated clearly the presence of seven hydrophobic transmembrane sequences[22]. The next major breakthrough in GPCR structural studies came 10 years later when two dimensional crystals of rhodopsin in phospholipid bilayers could be grown and analyzed by cryo-electron microscopy yielding a projection density map at a resolution of about 9 Å. These studies for the first time visualized unequivocally the bundle of seven transmembrane helices[129]. In 2000 the first high-quality three dimensional crystals of bovine rhodopsin in mixed micelles were obtained[130], allowing Palczewski *et al.* the determination of the first high-resolution structure of a GPCR at a 2.8 Å resolution[131] (figure 4a).

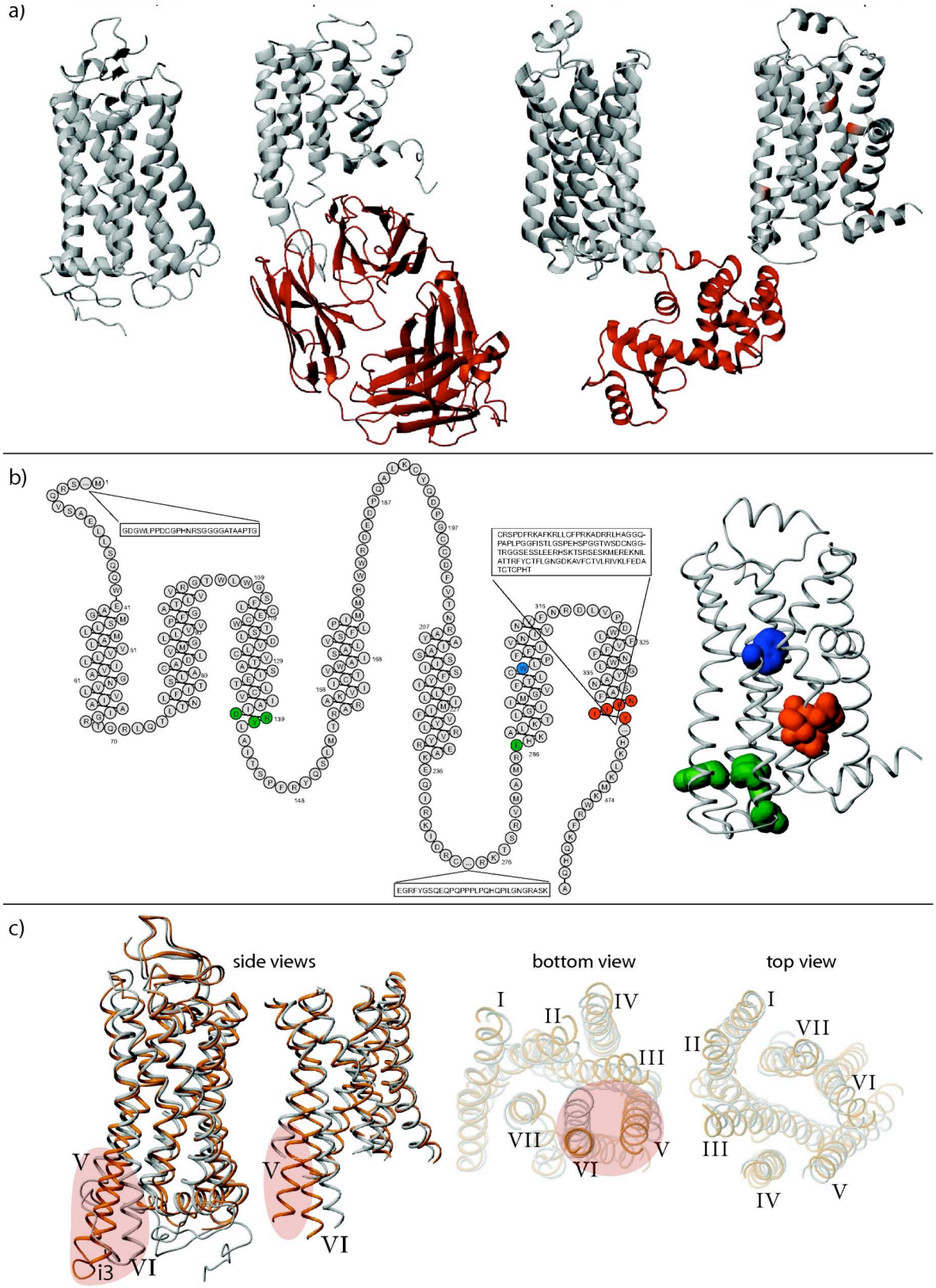


Figure 4: *a) Ribbon representations of three different GPCRs and the crystallizzation strategy which was applied. Rhodopsin was extracted from a native source, whereas the β-ARs were produced recombinantly in insect cells. Mutations introduced into β1-AR in order to obtain stable crystals and fusion (T4L) or interaction (Fab) partners for the crystallization of β2-AR are colored red. b) Snake plot representation of the β1-Ar with the ionic lock colored green, the rotamer toggle switch colored blue and the NPXXY motif colored red. These three motifs are also colored in the same manner in the crystal structure of β1-AR on the right. c) Overlay of the rhodopsin (gray) and opsin (orange) crystal structures. The main differences are in the i3 loop region and at the bottom of TM5 and 6. They are highlighted in red.*

The obtained structure essentially confirmed the predicted topology of GPCRs as consisting of an N-terminal extracellular domain, a cytoplasmic C-terminal domain and seven transmembrane helical segments (TM1-7) interconnected by three intracellular loops (ICL1-3) and three extracellular loops (ECL1-3), but also revealed the presence of a non-anticipated short antiparallel β-sheet in ECL2.

The crystal structures of four other GPCRs, namely the human β2-adrenergic receptor[132-136] (β2-AR), the avian β1-adrenergic receptor[137] (β1-AR), the human A_{2A} adenosine receptor[138], and bovine opsin[139], followed in 2007 and 2008. Except for opsin all of these structure represent "ligand bound" states: rhodopsin is covalently bound to retinal[131], the β1-adrenergic receptor is bound to the antagonist cyanopindolol[137], β2-adrenergic receptor to the inverse agonist carazolol[132-134] and timolol[135,136], and the A_{2A} adenosine receptor is in complex with the antagonist ZM241385[138].

Of these GPCRs rhodopsin is the only one which has not been expressed recombinantly, but could be extracted in sufficient amounts from natural sources (*vide supra*). A second major challenge in the handling of GPCRs is their low thermal stability[140]. Of the structures determined thus far, only rhodopsin didn't require any stabilization strategies to allow crystallization. The main strategies followed to allow successful crystallization of the above mentioned GPCRs shall be outlined briefly in the following paragraph. As was already mentioned rhodopsin could be extracted from a natural source in sufficient amount. Additionally, rhodopsin thus far has been the only GPCR with reasonable thermal stability. Therefore the crystallization of rhodopsin didn't require any recombinant expression strategies and/or mutagenesis screenings (figure 4a). This is much in contrast to the other GPCRs, for which crystal structures have been published. The β2-adrenergic receptor was truncated near the C-terminus and N-linked glycosylation was removed to increase the thermal stability. Additionally it carried an affinity tag for efficient purification. The low solvent exposed surface area necessitated the application of different strategies to increase the number of crystal lattice-forming contacts. In one determined structure β2-AR was co-crystallized with an antibody Fab fragment stabilizing the TM5/TM6 regions[132] (figure 4a). In the other determined structure the TM5/TM6 region was stabilized through the insertion of the T4 lysozyme sequence into the third intracellular loop (ICL3)[133,134] (figure 4a). Comparison of the structures of the two different constructs showed minimal deviations, indicating that the fusion of the T4 lysozyme sequence had little effect on the overall structure of β2-AR. The same strategy (truncation, removal of glycosylation, affinity tag and insertion of the T4 lysozyme sequence into ICL3) was chosen for obtaining high quality crystals of the human A_{2A} adenosine receptor[138]. The β-AR1 was truncated in the flexible ICL3 and N- and C-termini. Additionally the palmitoylation site was removed and an affinity tag introduced. A number of point mutations (7 in total; all in the TM helices) was introduced to increase the thermal stability of the receptor[137] (figure 4a). Except for the ICL2, which

comprises a short helical segment, the overall structure is very similar to the β2-AR. Furthermore, all GPCRs structures determined thus far, with the exception of opsin, were bound to a ligand, which stabilized their structure.

The sequence comparison of many GPCRs[141] and the recently published crystal structures (*vide supra*) have led to the postulation of key residues/motifs important for the structure and function of GPCRs. A number of highly conserved proline residues in the TM helices in TM4, TM5, TM6 and TM7 have been identified, which are thought to introduce kinks in the helices and thereby forming the ligand binding pocket[142]. Other conserved residues in the TM region of GPCRs have also been identified[141]. It could be shown experimentally, that several conserved residues have a strong influence on the active/inactive equilibrium of the receptors. One of them is a tryptophan residue in TM6, commonly embedded in a CWxP motif and named "rotamer toggle switch" (figure 4b) as the side chain rotameric conformation of the Trp plays a role in receptor activation. The residues which comprise this switch are coupled and modulate the bend angle of TM6 around the highly conserved proline in its center[143]. Another important motif is the (D/E)RY motif near the cytoplasmic end of TM3 (figure 4b). The strictly conserved arginine (residue 3.50 according to the Ballesteros-Weinstein numbering scheme [*vide infra*]) forms an ionic interaction with a conserved glutamate (6.30) on TM6, termed the "ionic lock". This "lock" has long been thought to keep the receptor in the inactive state[144-146]. Recent crystallographic data on the ionic lock, however, are not conclusive as to whether this is really the task of this motif[147,148]. The interaction pattern of an asparagine (7.49) and a tyrosine (7.53) both located in TM7 as part of a NPxxY motif (figure 4b) is also thought to play an important role in receptor activation[149]. A current model for the activation of GPCRs states, that as long as it is unliganded, the inactive state of a GPCR is stabilized by noncovalent interactions among different motifs such as the ones mentioned above. Structurally distinct ligands then lead to the breakage or retention of different combinations of these motifs[150]. These ligand-dependent response profiles are useful in the explanation of the different activation profiles, which can be observed within one family of ligands for a certain GPCR.

The solution of the β2-AR structure bound to the inverse agonist timolol[135,136] revealed the presence of a specific cholesterol binding site (cholesterol consensus motif, CCM) located between TM2, TM3 and TM4. Based on sequence similarities among multiple GPCRs in the CCM region, cholesterol binding might be a general feature of these receptors [135].

Presently eighteen rhodopsin and opsin crystal structures have been published (for a compilation of all GPCR crystal structures published so far see appendix E in chapter 4). Several of these, including the original publication[131], are of rhodopsin in the inactive ground state. But also for isorhodpsin (containing the non-native chromophore 9-*cis*-retinylidene)[151], two light activated intermediates populated early in the activation chain[152], a photoactivated receptor in the meta-I

state[153,154] as well as the opsin apo-protein[139] crystal structures have been published. Opsin is the retinal-free form of rhodopsin (the apo-protein), which is generated after photoactivation through the hydrolysis of the Schiff base.

All the rhodopsin structures discussed so far were determined on bovine rhodopsin. Additionally also the crystal structures of squid rhodopsin has been solved[155,156]. There are three main characteristics of light-activated rhodopsin: first, it shows an isomerization of the retinal chromophore from the 11-*cis* to the all-*trans* form, second, it manifests a spectral shift of its absorbance maximum from 500 nm to 378 nm, which is caused by the deprotonation of the Schiff base linking the chromophore to Lys296$^{7.43}$ (the numbering scheme used is the three-letter amino acid abbreviation followed by the residue number and the Ballesteros-Weinstein numbering of the residue in superscript) of the protein[157,158], and third, a hydrogen bond between Arg135$^{3.50}$ of the (D/E)RY motif in the third and Glu247$^{6.30}$ in the sixth transmembrane helix, the so-called "ionic lock", is broken[159]. The structure of the photoactivated rhodopsin shows all of these characteristics[154]. Comparison between the ground and activated structures reveals the largest differences in the third cytoplasmic loop of the receptor (figure 4c). This loop becomes completely disordered upon activation, with some residues missing in the electron density map. However, already in the ground state this loop is more disordered than other regions in the receptor as is evident from its increased temperature factor. The breaking of the ionic lock upon receptor activation is seen from an increase in the distance between Arg135$^{3.50}$ and Glu247$^{6.30}$ from 2.7 Å to 4.1 Å in the ground and activated states, respectively.

In the structures of the β1- and β2-adrenergic receptors, which are both in the inactive state, the ionic lock motif is broken. In the case of the β1-AR the inverse agonist cyanopindolol is bound[137], and it is not clear as to why a compound lowering the basal activity of the receptor should disrupt the ionic lock, a feature which is commonly attributed to the activation of the receptor. In the case of the β2-AR even less conclusion about the broken ionic lock can be drawn, because an arginine residue of the T4 lysozyme fusion, which has been inserted into the third intracellular loop to enhance protein crystallization, disrupts the (D/E)RY motif[133,134].

As might be expected, the most significant structural differences between the published receptor structures lie in the extracellular loops. For instance the second extracellular loop 2 (ECL2) of rhodopsin together with the N-terminus forms a short β-sheet, that shields the covalently bound bound retinal from bulk solvent, thereby preventing Schiff base hydrolysis. This shielding is further enhanced by the conformation, which is adopted by the glycosylated N-terminus[160,161]. The ECL2 of the two adrenoreceptors, on the other hand, contain both a short α-helix which is stabilized by intra- and inter-loop disulphide bonds. The N-terminal domain of these two receptors is disordered[133,134,137]. It appears that the extracellular face of the adrenoreceptors has, in contrast to

rhodopsin, evolved in such a way as to allow easy access to the ligand binding site.

The ECL2 of the A_{2A} receptor shows constrained conformations exposing the ligand-binding cavity to bulk solvent. The conformational constraints on the loop are mainly imposed by multiple disulphide bonds[138]. The advent of additional high-resolution structures will be needed to judge, whether the open binding pocket is a general feature of GPCRs that bind soluble small-molecule ligands.

The binding sites of retinal, cyanopindolol and carazolol are overlapping in superpositions of rhodopsin[131], β1-AR[137] and β2-AR[133,134], respectively. The binding orientation of the ligand is parallel to the membrane bilayer in all three above mentioned receptors, which is in contrast to the perpendicular orientation of ZM241385 bound to the A_{2A} receptor [138]. The aliphatic tail of carazolol in the β2-AR structure follows a very similar path to retinal in rhodopsin. In both cases the ligand extends from TM7, where it makes polar interactions in the case of carazolol and is covalently bound for retinal, to the interface of TM5/TM6. It is only here that the orientations of the two ligands start to differ as retinal extends deeper towards the center of the membrane, where it interacts with Trp265$^{6.48}$ of the "toggle switch".

The most interesting question concerning the activation is probably, how the changes in the interactions in the ligand binding pocket can be propagated to the cytoplasmic loops and thereby translated into G-protein activation. The structure of opsin highlights the transmembrane helix rearrangements which take place during receptor activation[139] and the structure of opsin in complex with a G-protein peptide can provide clues about the actual conformational state of the TM helices and the intracellular loops of a GPCR in the active state[162].

The opsin molecule doesn't correspond to activated rhodopsin, but rather represents a state later in the activation chain, and can be considered something like a desensitized or reduced activity form of rhodopsin. However, it does show residual activity in activation of transducin which is markedly higher as compared to ground state rhodopsin[163], for which basal activity is almost zero. Through infrared spectroscopy studies it was inferred, that opsin at low pH adopts a conformation that resembles the active state of metarhodopsin II[164,165]. The crystal structure of opsin was determined at low pH (5.6)[139] and several conformational changes of residues located near the ligand binding-pocket relative to rhodopsin could be observed. Most importantly the side chain of Trp265$^{6.48}$, the “toggle switch", moves into the space previously occupied by the ionone ring of the retinal. The most dramatic structural differences between rhodopsin and opsin are observed at the cytoplasmatic surface of the molecule. The cytoplasmatic ends of TM5 and TM6 and the interconnecting ICL3 are shifted away from the center of the helix bundle. At the same time the ends of the two helices move closer to each other. This movement observed in the crystal structures is consistent with previous biophysical data obtained by site directed spin labeling studies[166,167]. The ionic lock is broken and

Arg135$^{3.50}$ forms a new interaction with Tyr223$^{5.58}$, while Glu247$^{6.30}$forms a salt bridge with Lys231$^{5.66}$. Tyr306$^{7.53}$ from the NPXXY motif on TM7 occupies some space in opsin which has been occupied by TM6 in dark-state rhodopsin, thereby probably stabilizing the activated conformation characterized by the TM5/TM6 movement. Overall the change from inactive rhodopsin to an activated form is thought to create a cavity between TM3, TM5 and TM6 to which transducin can bind, a hypothesis which is also corroborated by the crystal structure of opsin in complex with a C-terminal peptide of transducin[162]. The transducin peptide adopts a helical conformation and interacts with opsin in an amphipathic manner. No significant structural changes are seen upon binding of opsin to the C-terminal transducin peptide.

All these findings, which highlight the structural changes required for the activation of rhodopsin, however, do not answer the question, how the binding event near the extracellular side of the receptor is translated into the structural changes observed on the cytoplasmic side. More biophysical data, especially such with time resolution, will be required to answer this question.

3.2.5 Ligand binding in class B GPCRs

With roughly 700 members class A GPCRs represents the largest subfamily of the 7TM receptors. Class B receptors with approximately 15 members are distinguished by their large N-terminal extracellular domain (ECD) of 100-160 residues which have a prominent role in ligand binding[1,168]. The ligands of class B GPCRs are endogenous peptide hormones such as glucagon, parathyroid hormone (PTH), corticotropin-releasing factor (CRF), growth-hormone releasing factor (GRF), pituitary adenylate cyclase (PACAP), glucose-dependent insulinotropic polypeptide (GIP), glucagon-like peptide 1 (GLP-1) and many more. Even though to date no high-resolution structure of a class B GPCR has been reported to date, the principles underlying ligand binding and receptor activation are better understood than for the class A GPCRs. This is mainly due to the fact that the large ECD1 of various receptors have been expressed recombinantly and could be used to perform ligand binding studies in vitro, leading to several high-resolution structures of free ECD and in complex with the cognate ligand. These include a subtype of human PACAP receptor[169], human PTH receptor[170], human GIP receptor[171], human GLP-1 receptor[172], the murine type-2β CRF receptor[173,174] and the human type-1 CRF receptor[175,176].

Receptor binding occurs generally through the C-terminal residues with the ligands adopting a helical conformation upon receptor binding, thereby generating the conformationally active state. Receptor activation is then thought to occur through an interaction of the structured N-terminal part with the transmembrane region of the receptor[177]. This model is consistent with all of the above structural studies and also explains why N-terminally truncated peptide ligands generally act as

antagonists.

All these ECDs share a common fold (called "secretin family recognition fold") comprised of an N-terminal helix, two antiparallel two-stranded β-sheets and three conserved disulfide bonds. Most notably the NMR studies on the ECDs of the two CRF receptors showed two unstructured loops undergoing dynamics on the millisecond to second timescale leading to line broadening effects. One of these loops gets structured upon ligand binding[173-175].

3.3 GPRC classification and residue nomenclature

Contrary to the structural homology common to all GPCRs, these receptors do not show any overall sequence homology across all members of the family[141,178]. However, significant sequence homology can be found within the three major subfamilies. Family A, which is the largest subfamily, is represented by rhodopsin but also includes the β-adrenergic receptors. Family B receptors are related to the glucagon receptor and receptors of family C to the metabotropic neurotransmitter receptors. Two additional small subfamilies D and E are made up by the yeast pheromone receptors STE2 and STE3, respectively. And in the organism *Dictyostelium discoideum* a sixth subfamily of cAMP receptor constitutes subfamily F[178].

Family A is by far the largest and most studied. It can be subdivided into smaller subgroups. Even though the overall homology among all family A receptors is low there are a limited number of highly conserved key residues. This high degree of conservation indicates that they have an essential role for the structural or the functional properties of the receptor. The only residue which is absolutely conserved among all family A GPCRs is the arginine in the E/DRY motif near the cytoplasmic side of TM3 (*vide supra*)[141,178]. An additional characteristic for family A GPCRs is a conserved disulfide bridge between ECL1 and ECL2 and several conserved residues within the TM helices. Within the family A GPCRs, however, differences regarding the binding site of the ligand exist. It is these differences which serve as the basis for a further division of this family into three different subgroups: Subgroups 1 and 2 differ from subgroup 3 by their short N-terminal domains (10-50 residues vs ~100 residues, respectively). The distinguishing feature between subgroups 1 and 2 is the location of the ligand binding pocket on the receptor. Whereas for subgroup 1 the ligand binding pocket is buried in the receptor's TM region, the binding pocket for subgroup 2 members is formed by the extracellular parts (N-terminus and loops) of the receptor. Representatives of subgroup 3 are thought to bind their ligands mainly via their N-terminus[179] (figure 5).

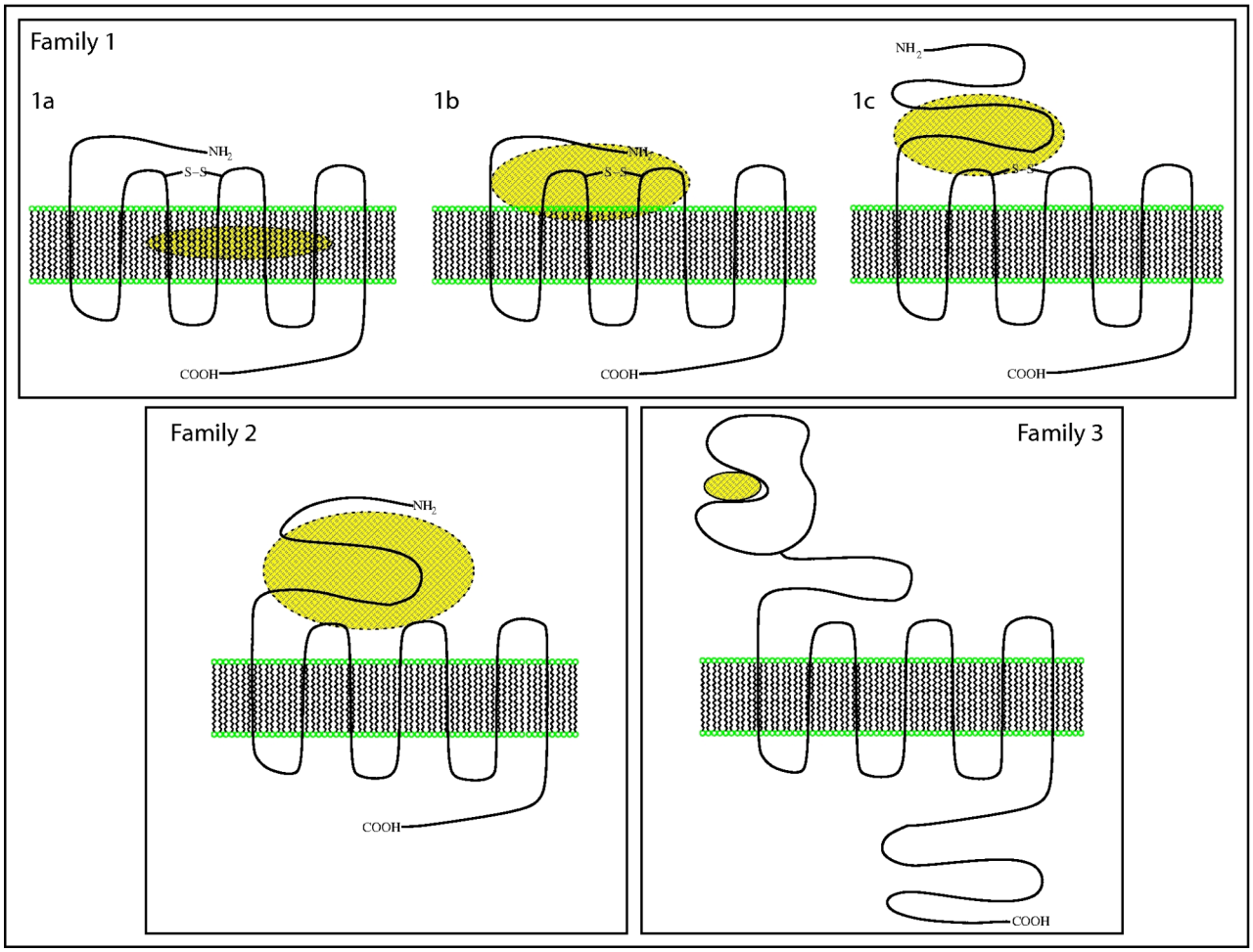


Figure 5: *Classification of GPCRs into three main families according to structural characteristics as explained in the text.*

Family B of GPCRs is made up of approximately 20 different members, which bind a variety of peptide hormones. Importantly the (D/E)RY motif is absent in this GPCR subfamily, which has a large (~100 residues) extracellular N-terminal domain containing several cysteines, which presumably form a disulfide bridge network[180], as the most characteristic structural feature. The characteristic disulfide bridge connecting ECL1 and ECL2 is absent in family B GPCRs[179].

The N-terminus of family C GPCRs is even longer, ranging from 500 to 600 amino acids and sharing sequence similarity with periplasmic binding proteins of bacteria[181]. The ligand binding site is believed to be located exclusively in this large extracellular domain[182].

Because of the large variation in the length of the GPCR polypeptide chain between different subfamilies, but also within such subfamilies, a residue numbering system not primarily based on the position in the primary sequence is required to allow for the comparison of different GPCRs. Different numbering schemes have been devised[183-185], with the one from Ballesteros and Weinstein being the most commonly used nowadays[185]. In their scheme each helix is labeled with its number (i.e. from 1 to 7). The most conserved residue in each helix is then arbitrarily given the number 50 (see figure 8 for an example). Residues towards the C-terminus will sequentially be given numbers

bigger than 50, and residues towards the N-terminus will be labeled with numbers smaller than 50. This residue number will be separated from the helix number by a dot and both numbers are in superscript, if they are added to the three-letter abbreviation code of an amino acid. As an example the most conserved residue in TM5 of rhodopsin is a proline. This proline would therefore be labeled as Pro$^{5.50}$. Its preceding isoleucine would be called Ile$^{5.49}$ and its succeeding leucine Leu$^{5.51}$.

3.4 Alternative strategies for the structural characterization of GPCRs

Even though high resolution crystal structures of a few GPCRs could be obtained in the last years, the efforts leading to such a structure are still huge, mainly because crystallization of membrane proteins is still a more than challenging task. Since these proteins contain large areas of hydrophobic surfaces, they must be isolated in detergents, which "mask" a large proportion of the accessible surface area and render it incapable of forming protein-protein contacts, which are essential for forming three-dimensional crystals. The second major technique allowing the determination of high-resolution protein structures is NMR. While this approach has become very successful in the structure determination of soluble proteins, its use to study membrane proteins has been much more limited. The reasons for this are manyfold. Obtaining suitable amounts of membrane proteins for structural studies is associated with many difficulties. While issues concerning the low levels of expression, the tedious purification process, and the often inherently low thermal stability of membrane proteins also concern crystallographic studies, an additional factor prevents the successful application of NMR methods to the study of many membrane proteins. This is the molecular weight (MW) of the proteins under study. As there is no theoretical MW limit for the study of a system by X-ray crystallography crystal structures of systems as large as the (entire) ribosome[186-188] or the fatty acid synthetase[189,190] have been determined (both with molecular masses of ~2.5 MDa). Major technological improvements of the NMR hardware include the increasing magnetic field strengths[191], the design of cryogenically cooled probes[192] and the development of microcoil NMR[193]. On the methodological side the MW limit for structure determination has been pushed from 10 kDa in the early days of NMR[194], where only proton-proton correlation experiments could be recorded, to ~50 kDa through the introduction of triple resonance techniques[195], to currently ~100 kDa through the development of elaborate labeling schemes[196,197], perdeuteration and transverse relaxation optimized NMR spectroscopy[198,199]. Currently the largest molecules whose structures have been solved by NMR completely are around 80 kDa. One example is the 82 kDa malate synthase G from *E. coli*[200]. While the polypeptide chain of a typical GPCR alone is usually in the 50 kDa range, the detergent molecules associated with the polypeptide chain make the effective MW much larger.

4. NMR spectroscopy with membrane proteins

4.1 Introduction

Even though membrane proteins make up roughly 50% of the proteome of an organism, structural information on this class is much underrepresented compared to soluble proteins. Of the membrane proteins of known structure those with large extramembraneous domains are predominant, while information on small membrane proteins with a high membrane-embedded portion is scarce. This imbalance stems from the requirement of protein-protein contact areas for the growth of three-dimensional protein crystals[201]. From an NMR point of view exactly these small membrane proteins are ideal candidates for structure determination. It is thus likely that solution NMR will play an important role in delivering structural information on this class of proteins.

4.2 The role of detergents in NMR spectroscopy with membrane proteins

As was already mentioned membrane proteins must be solubilized from their native lipid environment through detergents. In order to make a membrane protein amenable to solution NMR spectroscopy, the chosen detergent must keep the membrane protein in its native conformation and additionally the detergent-solubilized membrane protein must be small enough to allow fast isotropic tumbling in solution. Detergent micelles are the most commonly used solubilizing agent for solution NMR studies. A series of zwitterionic and anionic detergents is available and have been used successfully in NMR spectroscopy[202], whereas uncharged detergents are less frequently used for NMR studies of membrane proteins. Aside from the charge of their head-groups, detergents are characterized by their critical micelle concentration (cmc), monomeric molecular weight and aggregation number. From the latter two the molecular weight of a micelle aggregate can be calculated. Even though extensive screening studies have been carried out in order to find a universal detergent for NMR studies, generally the optimal solubilization conditions, including the choice of the detergent, must be worked out for each membrane protein[203-205]. For enzymes native structure can be tested through functional assays as demonstrated for the diacylglycerol kinase (DAGK)[206-209] or the β-barrel bacterial outer membrane protein PagP[210-212].

Even though detergents such as sodium dodecylsulfate (SDS)[213], dodecylphosphocholine (DPC)[214] and lauryldimethylamine N-oxide (LDAO)[215], which have the charged head group directly attached to the alkyl chain, have been used successfully in structural studies (see references for examples), this family of detergents has a propensity to destabilize proteins[202]. More promising for the

preservation of native structure and biological function are detergents, which carry a polar, but uncharged spacer between the charged head group and the apolar tail. Two prominent members of this family are dihexanoyl phosphatidylcholine (DHPC)[216] and the lysophospholipids[204] with alkyl chain lengths of 14 and 16 C-atoms and different headgroups such as phosphatidylcholine or phosphatidylglycerol.

Micelles are usually depicted as spherical aggregates. In nature, however, they often assume oblate or prolate shapes[217-219]. The molecular weight of a typical micelle is in the range of 50-100 kDa. When a membrane protein is solubilized in a detergent micelle, a so called "mixed micelle" is formed. In case of large membrane proteins with molecular weights in the range of the micelles the molecular weights of micelle and protein are usually not additive. Rather the hydrophobic surface area of the protein determines how many detergent molecules can be bound, and hence ultimately determines the size of the mixed micelle[206]. Therefore the size of the protein-micelle complex cannot simply be minimized just by choosing a detergent type known to form small micelles.

Apart from micelles small isotropic bicelles, composed of long-chain lipids forming a planar mini-bilayer, whose rims are covered by short-chain lipids, have been used successfully in solution NMR. Due to their planar surface they represent the most membrane-like environment in which solution NMR is still possible. Bicelles have also been used extensively in solid-state NMR studies of membrane proteins[220-224] and are therefore an ideal system for combined solution/solid-state NMR studies. Apart from membrane associated peptides[220] also a few membrane proteins studied successfully in bicelles by solution NMR. Among them are the β-barrel proteins OmpA[225] and OmpX[13]. A further development of the bicelle system are lipid-protein nanoscale bilayers[226], which may prove useful for solution NMR in the future as indicate promising results with the mitochondrial voltage dependent anion channel VDAC-1[227]. They are reconstituted high density lipoprotein particles composed of a planar patch of ~160 lipid molecules, with the rim of the patch covered by the lipid binding protein apolipoprotein A-I.

Two other promising solubilization media for membrane proteins are amphipols[228] and reversed micelles[229]. The latter have been used to enclose soluble proteins within their aqueous interiors. Dissolving these large complexes in low viscosity alkanes such as pentane or hexane results in more rapid overall tumbling than can be observed for the enclosed soluble protein in water[230-232]. Applications of this system for the study of membrane proteins are currently in progress[233].

4.3 Isotope labeling strategies in NMR spectroscopy with membrane proteins

Obtaining chemical shift assignments of membrane proteins requires ^{2}H, ^{13}C, ^{15}N-labeled samples in almost all cases. Fortunately for many of the screening processes (refolding efficiency, sample

stability, spectra quality) the much cheaper ^{15}N-enrichment alone is sufficient. Generally ^{15}N-^{1}H heteronuclear single quantum correlation (HSQC) spectra can be recorded on solely ^{15}N-labeled samples of up to 100 kDa in size[234]. As most NMR experiments rely on the excitation and detection of ^{1}H magnetization, deuterated proteins necessitate the introduction of protons for ^{1}H-based experiments to work. This is usually done by dissolving the denatured protein in H_2O, which leads to the exchange of ^{2}H to ^{1}H at all exchangeable sites including the backbone ^{15}N-^{1}H amide group. In addition to protonation at the backbone amide sites in an otherwise highly deuterated background, its is also possible to selectively protonate methyl groups of Val, Leu and the δ1 of Ile by supplementing an *E. coli* minimal medium with precursors of these amino acids, which carry only protons at the methyl sites and deuterons in all other sites[197]. The precursors which allow this type of labeling are α-ketoisovalerate and α-ketobutyrate[70,235]. Provided a complete backbone assignment is available the methyl ^{13}C and ^{1}H chemical shifts can be assigned by relaying proton magnetization generated on the methyl to the backbone N-H groups by experiments specifically designed for highly deuterated proteins carrying selectively labeled methyl groups[236,237]. Recently also the selective methyl protonation of alanines became feasible at reasonable cost[238]. Ala is, in contrast to Val, Ile and Leu, evenly distributed across a protein's sequence and therefore is a very attractive residue for the selective labeling. In a highly deuterated protein the lack of protons other than the ones in the exchangeable sites hampers the structure determination based on proton-proton NOEs. The selectively protonated methyl groups therefore serve as an important source of additional NOE based structural restraints[239], because generally NH-NH NOEs alone are insufficient to determine the global fold of a protein[236]. An exception are the β-barrel proteins in which NH-NH NOEs give many long-range restraints, allowing to accurately define the global fold[234].

4.4 Alternatives to traditional NOE based structural restraints

4.4.1 Residual dipolar couplings

The scarcity of NOE restraints in highly deuterated proteins necessitates the access to other structural information, in order to obtain accurate protein structures. Residual dipolar couplings induced by partial alignment of the protein in the magnetic field is able to deliver such information[240,241]. Alignment is achieved by special media of significant magnetic anisotropy that hence have a preferred orientation in the magnetic field. The non-isotropic orientation of these media is then (partially) transferred to the protein under study. Commonly used alignment media include large bicelles[242], bacteriophages[243,244], mechanically strained polyacrylamide gels[245,246] and DNA nanotubes[247,248]. Unfortunately the former two are incompatible with the detergents required

for the study of membrane proteins[249]. Strained polyacrylamide gels have been successfully used to align membrane proteins[250,251], but often decrease the tumbling rate, especially in large systems. The DNA nanotubes seem to be a promising system for aligning membrane proteins, but relatively little knowledge is available so far, which is also due to their complicated synthesis.

4.4.2 Paramagnetic relaxation enhancement

Unpaired electrons of paramagnetic molecules can lead to a dramatic increase in relaxation rates of nuclear coherences through electron-nuclear spin dipolar interaction, because of the large magnetogyric ratio of the electron ($\gamma_{2H}/\gamma_{1H} \sim 660$)[252]. This phenomenon is called paramagnetic relaxation enhancement (PRE). Paramagnetic compounds in the solvent, such as dissolved O_2, can lead to an overall broadening of the resonances of the molecule under study. In case a paramagnetic species, such as a metal ion or a nitroxide spin label, is attached to a protein, despite the presence of broadened lines in vicinity of the paramagnetic center, useful structural information can be obtained. Whereas NOE restraints are limited to ~6 Å, PRE is active over considerably larger distances of up to 25-35 Å[253,254] and has hence been recognized early as a potentially important effect for obtaining long-range distance information[255]. In proteins where NOE restraints are difficult to obtain, PRE restraints can thus be beneficial for defining the global fold. An additional advantage of PRE based distance restraints is, that they can be obtained much more easily than NOE restraints, especially for large proteins where signal overlap and poor resolution often hamper the acquisition of NOE restraints, and they do not necessarily require assignment of sidechain resonances. In addition, they can be placed in more flexible regions, which are difficult to assign. Only certain metalloproteins naturally possess paramagnetic centers. Therefore this class of molecules is the one most extensively studied making use of PRE[256-258]. Proteins lacking a natural paramagnetic center can often be labeled with paramagnetic nitroxide centers through site-directed spin labeling (SDSL)[259]. Application of this technique has been used to tackle lacking long-range NOE distance restraints such as unfolded[260]/disordered[261-265] proteins, but also large soluble proteins[253] and most notably membrane proteins[209,266,267]. In the case of the integral membrane proteins Mistic[267] from *B. subtilis* and diacylglycerolkinase (DAGK)[209] from *E. coli* PREs were used in addition other structural restraints. In the case of the *E. coli* outer membrane protein A[266] PREs from SDSL on 11 single cysteine mutants were used to refine the previously published solution structure. However, it could also be shown in that case that the PREs alone yielded a structure of reasonable quality.

4.5 New experimental frontiers in membrane protein NMR spectroscopy

As traditional proton based NMR reaches its limits in the study of large systems, alternative strategies, such as NMR experiments relying solely on ^{13}C have been developed[268,269].

The larger a protein, is the slower it tumbles in solution, which implicates fast relaxation of coherences and concomitantly broad lines in the NMR spectrum. At high magnetic field the dipole-dipole interaction and the chemical shift anisotropy (CSA) are the dominant relaxation pathways. In an scalar coupled I-S spin system with the S spin having a large CSA (as is for instance the case for the ^{1}H-^{15}N spin system of the backbone amide) the IS correlation spectrum is composed of four components of differing line widths. In the narrowest of these components the dipole-dipole interaction and the CSA partially compensate each other. The TROSY (transverse relaxation optimized spectroscopy) experiment is designed such that only the component with the narrowest line width is retained in the spectrum and all the other components are discarded[198]. The original TROSY experiment has been enhanced[270,271] and its principle has been introduced in a variety of multidimensional experiments, enabling resonance assignment[199,272-274] and structure determination[275] exploiting the TROSY phenomenon. The introduction of the TROSY principle pushed the molecular weight limit for solution NMR to ~100 kDa. Other relaxation interference mechanism can also be exploited, giving rise to TROSY phenomena for groups which don't show significant CSA such as the methylene and methyl groups[276-278].

4.6 GPCRs studied by NMR spectroscopy

The application of solution NMR to the study of GPCRs can be divided into four main fields: First the characterization of GPCR fragments, second the examination of ligands bound to GPCRs[279-282], third the detection of selectively labeled probes in rhodopsin[283-285]. In the fourth field, the study of full-length GPCRs only few successes could be achieved so far[286].

4.6.1 Detection of selectively labeled probes in GPCRs

Several different NMR observable probes have been incorporated into rhodopsin and could be used for its study by solid-state and liquid-state NMR methods.

The incorporation of a selectively ^{15}N- or ^{13}C-labeled amino acid types into rhodopsin has been one of the main strategies to study full length GPCRs by NMR. Such selective labeling is achieved by adding appropriately labeled amino acids or their precursors to the growth medium of mammalian cells (HEK cells) used for the recombinant production of rhodopsin. In an initial study[283] lysine carrying a ^{15}N-label at its α-amino group ([α-^{15}N]lysine) was included in the growth medium of the HEK cells stably expressing rhodopsin at about 2.5 mg/l of cell culture[287]. 11 lysine residues are

distributed over the rhodopsin sequence, 9 of which are in the cytoplasmic, 1 in the intradiscal (extracellular), and 1 in the TM domain (Lys296, to which the retinal chromophore is attached). Samples of native rhodopsin (as judged by a A_{280} to A_{500} ratio of ~1.6[288]) in dodecyl maltoside (DM; 4-7%) or octyl glucoside (OG; 1.4%) micelles measured at 4 °C showed only one peak in the [^{15}N,^{1}H]-HSQC spectra. Only going to higher temperatures and/or including 5% SDS in the samples led to the appearance of additional peaks, however with concomitant denaturation of rhodopsin. Based on an HNCO experiment[289], a protease digestion assay and antibody binding studies the visible peak under native conditions was assigned to the lysine residue located in the C-terminus of rhodopsin. As the size of the rhodopsin containing micelles was estimated by dynamic light scattering (DLS) to be around 100 kDa, the lack of the remaining 10 signals was attributed to conformational exchange processes on the micro- to millisecond time scale leading to exchange broadening of the respective resonances and not to a total molecular weight of the system outside the accessible range of such basic 2-dimensional experiments as the TROSY version of the [^{15}N,^{1}H]-HSQC. In a following study conducted by the same group tryptophan carrying a ^{15}N-label at its α- and ε-nitrogens ([α,ε-^{15}N]tryptophan) was included in the growth medium of the same HEK cell line used in the previous study[287]. Four of the five tryptophans of rhodopsin are located in the TM region and only one is in the extracellular domain. The [^{15}N,^{1}H]-HSQC spectra of rhodopsin showed more than the expected 10 peaks. Whereas the expected 5 from the side chains are observed in the characteristic indole region more than 5 peaks are seen in the backbone amide region. This led to the conclusion that the backbone showed conformational exchange in the micro- to millisecond time scale, whereas the side chains had less conformational freedom. This was supported by inspection of the packing density around the Trp side chains in the rhodopsin crystal structure. This conclusion is opposite to what is commonly observed in proteins, where more often the side chains show conformational flexibility rather than the backbone. The conformational exchange processes observed in the backbone in both studies with rhodopsin might be a general feature of membrane proteins, as it was also observed in other families such as the β-barrel OmpA[214,290], bacteriorhodopsin[291,292] and diacylglycerol kinase[209,293].

The incorporation of selectively labeled amino acids into an otherwise unlabeled background can mostly be achieved *in vivo* in *E. coli*[294] and also in eukaryotic cells such as Sf9 insect cells[295]. This *in vivo* approach assures that it occurs under conditions which will not disturb the tertiary structure of the protein, which is certainly an advantage of this technique. The disadvantage of this technique is that the number of accessible labels is limited to the 20 natural amino acids. Even though in the last years significant progress has been made in the design of bacterial strains which allow the incorporation of “unnatural” amino acids into proteins[296,297], the application of these strategies is still not trivial. Furthermore they are not applicable to any desired label. Therefore the *in vitro*

introduction of labels is a widespread strategy. While this approach theoretically allows the incorporation of almost any conceivable label, it is limited by the availability of a suitable coupling method and on sterical limitations imposed by the protein-label pair. Since any label needs to be coupled to a reactive group of the polypeptide chain, often also selectivity issues arise. A commonly applied strategy to overcome this problem is the generation of (single-)cysteine mutants of a protein and coupling the label of choice via a thiol-reactive group to the mutant protein. A particularly interesting label for NMR studies is ^{19}F, due to its high magnetogyric ratio and the absence of a ^{19}F background in proteins[298]. In a series of studies conducted in the Khorana lab[285,299], ^{19}F labels were introduced into rhodopsin through oxidative coupling of [2-tri^{19}F]trifluoroethylthiol to the side chains of native and mutated cysteine residues in the rhodopsin sequence. Changes in the positions and line-widths of the ^{19}F-resonances of a rhodopsin sample measured in the dark were observed after illumination of the sample[299]. Furthermore NOEs between ^{19}F-pairs allowed estimation of inter-residual distances and their comparison with previously published values determined from other experiments[285].

4.6.2 Examination of ligands bound to GPCRs

The CC chemokine receptor 5 (CCR5) plays a role in the regulation of immune cell trafficking and is activated by a number of endogenous chemokines[300]. It is also associated with HIV infecting host cells[301]. The structure of the CCR5 has been modeled based on the crystal structure of rhodopsin[302] and the structure of the N-terminal 15 residues has been recently solved in complex with the HIV envelope protein gp120 (*vide supra*)[303]. Furthermore the binding sites of the CCR5 N-terminus have been mapped onto the CCR5 binding chemokine RANTES[304]. To enable further biophysical studies of the interaction of CCR5 with its chemokine interaction partners, an insect cell expression system and a purification protocol for CCR5 have been established[282]. With this system sufficient quantities CCR5 could be purified to allow the characterization of its interaction with RANTES by isothermal titration calorimetry. The ^{1}H-spectrum of the CCR5 in DPC micelles showed good signal dispersion[282], encouraging further studies of CCR5 by NMR. In a previous study on the same system the interaction of RANTES with receptor fragments comprising the N-terminus and the three extracellular loops has been examined. The binding surface of RANTES was found to be strongly positively charged and to overlap with sequences which have been known to be important for blocking HIV infection[304].

On the cytoplasmic side the interaction partners of GPCRs are the G-proteins. Especially the third intracellular loop and the C-terminus of the GPCR are thought to be important for this interaction[305-308]. The precise mechanism by which an activated receptor molecule (R*) binds to the

G-protein and catalyzes nucleotide exchange remains unknown. However, several important regions for receptor binding within G_α and $G_{\beta\gamma}$ have been identified[309-311]. In G_α one of the best studied interaction sites is the C-terminus, due to its susceptibility to ADP-ribosylation by pertussis toxin[312]. A synthetic undecapeptide from the C-terminus of transducin was shown to be a competitive inhibitor of G-protein activation by light activated rhodopsin[34]. Structural changes in this peptide upon association with purified light activated rhodopsin were elucidated by transferred nuclear Overhauser effect (TRNOE) spectroscopy[313], revealing that the peptide gets structured only after light activation of rhodopsin[281]. When the NMR sample consisted of the peptide and dark adapted rhodopsin, no long-range interactions could be detected in the NOESY spectrum, whereas after exposure to light a number of long-range interactions could be detected, indicating the assumption of a defined structure of the peptide[281].

Another ligand conformation determination was carried out for the pituitary adenylate cyclase activating polypeptide (PACAP) bound to its receptor, the PACAP receptor[280]. In their study Inooka *et al.* have determined the receptor-bound conformation of a truncated variant of the 27 residue PACAP. This truncated variant was the first 21 residues of the full-length PACAP and was therefore called PACAP21. The truncated PACAP21 was chosen, because it has a higher dissociation rate and weaker affinity towards the receptor, which puts PACAP21 in the fast exchange regime on the NMR chemical shift timescale. In the NOESY spectrum of a sample consisting of PACAP21 in presence of the receptor both TRNOE and NOE peaks were observed. While the former arise from the receptor-bound form, the latter stem from free PACAP21. Displacement of PACAP21 from the receptor was achieved by addition of the higher-affinity PACAP, allowing to measure the NOESY spectrum of PACAP21 free in solution. Subtraction of the two NOESY spectra yielded a spectrum showing only TRNOE peaks, from which the conformation of the receptor-bound PACAP21 could be calculated. Except for the N- and C-terminus all residues showed defined backbone and side chain conformations. The first 3 residues were in a disordered extended conformation and were followed by 5 residues comprising two unusual β-turn structures. The structured remainder of the peptide was α-helical. This receptor-bound conformation was compared with the structure of full-length PACAP bound to DPC micelles. Except for the β-turn region the two structures showed striking similarities, suggesting that the helical surface of the peptide interacts with the micelles and the receptor in a similar manner.

The muscarinic acetylcholine receptors (mAChRs) belong to the GPCR family. For the related nicotinic acetylcholine receptor (nAChR), which does not belong to the GPCR family, the conformation of the bound acetylcholine ligand has been determined by NMR in 1988[314]. In 2002 the conformations of the two mAChR ligands (S)-metacholine and (2S,4R,5S)-muscarine bound to

their receptor were determined by NMR[279]. First the conformation of the two ligands free in solution was determined from the coupling constants determined in one-dimensional ^{1}H-experiments. The conformation of the ligands bound to the mAChR was then determined from transferred NOE peak intensities obtained through subtraction of a NOESY spectrum of the respective ligand in complex with the mAChR and a NOESY spectrum of the same sample after addition of atropine, a strong mAChR antagonist, displacing the metacholine or muscarine ligands from the receptor. Rotations around the bond connecting C1 and C2 of the ligands were observed upon binding to the receptor supporting the binding of the ligands in a gauche conformation of the N-C1-C2-O dihedral angle. This was in contrast to earlier studies that reported the anti rotamers to be the receptor activating conformations[315]. The discrepancy between the two findings was explained by inaccuracies in the definition of the conformations in the earlier studies.

The ligand of rhodopsin was also derivatized with NMR-observable reporter groups and effects of photoactivation were observed on ^{19}F resonances in liquid-state NMR[316] and on ^{13}C- and ^{15}N-resonances by solid state NMR[317-325]. For these experiments rhodopsin was purified from bovine retina according to standard procedures, followed by an exchange of the natural retinal by the synthetic derivatives, by treating the purified rhodopsin samples with hydroxylamine (H_2NOH) under light irradiation. The resulting "apo" form of rhodopsin was purified by centrifugation and treated with the synthetic retinal derivative[326]. Both the liquid and solid state NMR studies sought to determine the conformation of the chromophore when bound in the 11-*cis* form in ground state rhodopsin and the all-*trans* form in light activated metarhodopsin II (MII).

A number of other ligand conformations when bound to their GPCR receptors has been studied by solid state NMR, including the histamine[327] and the neurotensin receptors[328].

4.6.3 NMR studies of full-length GPCRs

The complete resonance assignment of a protein gives the spectroscopist the possibility to probe a wealth of information with atomic resolution. Whereas the classical information sought in biological NMR is internuclear distances (mainly inter proton distances) defining the three-dimensional structure of the molecule under study, the measurement of other parameters, such as the dynamical properties of a nucleus, can yield just as important clues in elucidating biological processes. NMR is unique in being able to provide a wealth of structural and dynamics information on an atomic resolution. Therefore it is not surprising that the complete resonance assignment of a GPCR would be a milestone in the structural biology of this class of receptors. It should be stated here, that presently no such complete resonance assignment, let alone a complete NMR structure, of a GPCR has been obtained so far. Nevertheless a few promising steps towards achieving this goal

have been made. The results coming closest to this goal is were published in 2008 by Gautier *et al.*[286]. Even though their system is strictly speaking not a GPCR, it is also a 7 TM protein and can therefore be considered a "GPCR analog". They have reported the near complete backbone assignment, secondary structure determination, and analysis of backbone dynamics of sensory rhodopsin pSRII from *Natronomonas pharaonis*. pSRII belongs to the family of microbial rhodopsins and acts as a phototaxis receptor[329]. Both of these functions are driven by the *cis-trans* isomerization of the bound retinal molecule.

The success of this challenging project was dependent on optimal conditions in many respects. On the biological side it was clear that the study of such a large system necessitated a high-level of deuteration to meet the required resolution and sensitivity needs. These premises restricted the available host systems to *E. coli* and *P. pastoris* (*vide supra*). After purification of native pSRII in DM micelles the protein was exchanged into diheptanoylphosphatidylcholine micelles, which, at 50 °C, allowed the acquisition of high quality NMR spectra. On the NMR experimental side sufficient sensitivity and resolution could only be obtained in a reasonable amount of measuring time by the application of nonuniform sampling schemes in combination with maximum entropy reconstruction[330,331]. The uniform $^{13}C,^{15}N$-labeling scheme in conjunction with >95% deuteration allowed the assignment of almost all C_α, C_β and C' shifts through the use of TROSY[198] versions of 3D out-and-back HNCA/HN(CO)CA, HN(CA)CB/HN(COCA)CB, and HN(CA)CO/HNCO[199,332] experiments. After assignment the secondary structure was probed by calculating $\Delta\delta C_\alpha - \Delta\delta C_\beta$, a value which indicates α-helical secondary structure if positive and β-sheet structure if negative[333]. The observed pattern for this value was in very good agreement with the crystal structure[334]. The dynamics of the pSRII backbone were determined through measurement of the longitudinal (T_1) and transverse (T_2) relaxation times of the amide ^{15}N, and the $^{15}N^{1H}$-heteronuclear NOE. From the measured T_1 and T_2 values a rotational correlation time of ~20 ns was calculated, corresponding to a protein-detergent complex of 50-70 kDa. Some minor conflict with results from a previously published solid state NMR study of pSRII[335] was explained by the different hydrophobic environments used in two studies.

The solution NMR studies published on full-length GPCRs so far don't come close to the success achieved with sensory rhodopsin. In 2005 a publication reported obtaining high quality NMR spectra for uniformly labeled human vasopressin V2 receptor (V2R) in lyso-myristoylphosphatidylcholine (LMPC) micelles[336]. The initial promising results had to be, at least partially, revoked soon after, when it was realized that a majority of the observed peaks were in fact from a contaminant and not from the V2R[337]. After removal of the contaminant the spectra showed 80 of the total of ~350 expected peaks. Considering the scarcity of NMR data on full-length GPCRs this should still be considered a successful application of NMR to the study of GPCRs and a

promising starting point for the improvement of the system under study.

No solution NMR studies of uniformly labeled rhodopsin have been published so far. But a HEK cell line stably transfected with the opsin gene was used to label half of the rhodopsin residues[338]. [^{15}N,^{1}H]-HSQC spectra showed only around 20 peaks, which were identified as belonging to the C-terminus. Comparison with spectra of a peptide consisting of the 19 C-terminal residues of rhodopsin led to the conclusion, that the C-terminus of the full-length rhodopsin was unstructured under the chosen experimental conditions and hence gave rise to narrow line widths.

Decreased molecular tumbling rates due to increased molecular size limit the applicability of solution NMR. Solid-state NMR (ssNMR) under conditions of magic angle spinning (MAS) [339] doesn't suffer from these effects and has been shown to be able to characterize molecular structure and dynamics[340]. MAS has allowed to study the topology of various membrane proteins using a variety of different nuclei[340-342].

On the subject of GPCRs ssNMR has been most extensively used for the study of rhodopsin[322,324,343] - and its retinal chromophore[344]. Recent progress in the heterologous expression and reconstitution into lipid membranes of GPCRs has allowed to target other systems. In terms of ligand structure this includes most notably the determination of the receptor bound structure of human neurotensin fragment 8-13[327]. Even though not a GPCR, sensory rhodopsin II from *Natronomonas pharaonis* (NpSRII) is a seven-helix membrane protein containing a bound retinal chromophore and hence bears similarity in many respects with rhodopsin. The characterization of the secondary structure, dynamics, and topology of this system[335] points out the feasibility of ssNMR structural studies of heterologous 7TM proteins. More recently, other GPCRs such as the human Y2-receptor[345] or H1-receptor[328] could be reconstituted into environments suitable for ssNMR studies. Tapaneeyakorn *et al.* have recently given an excellent review over the solid state (and solution) NMR work carried out on the subject of GPCRs[346].

4.6.4 Alternative strategies for gaining structural and functional information of GPCRs

After the summary of NMR approaches to the study of GPCRs it seems clear that no straight forward techniques for these class of receptors exists so far. Therefore, alternative approaches to GPCR structure are desirable. One such approach is the study of suitable fragments of a receptor. By this approach the size of the system under study can be chosen by the experimenter. Evidently such a choice should follow a certain rationale. One such rationale is to express fragments corresponding to individual domains of a larger molecule. Of course the validity of such an approach needs to be questioned concerning its usefulness/value. It seems clear that this strategy is only able to allow limited conclusions as compared to the study of full length receptors.

Nevertheless there is evidence that peptides often retain their secondary structures when removed from their tertiary structure environment of a protein[347,348]. This finding can be expected to be true especially for secondary structural elements which are defined by short-range interactions as is the case for α-helices[349] and turns[350]. Additionally the design of so-called split receptors of rhodopsin[351] and the *S. cerevisiae* α-factor receptor Ste2p[352] has shown that receptor fragments can complement each other restoring the functionality of the full length receptor. Thus it seems to be a promising approach to elucidate the secondary structure of a helical bundle protein by studying fragments thereof. Many fragments of GPCRs have been studied with biochemical and biophysical methods. The finding that fragments derived from the cytoplasmatic face of rhodopsin were able to competitively inhibit the interaction between rhodopsin and its G-protein transducin[353] encouraged the first structure determination of such a fragment by NMR[354]. The biophysical characterization of such fragments often included structure determination by NMR or sometimes also X-ray crystallography. The experimental conditions are chosen such as to mimic the native environment encountered in biological membranes. Often organic solvents or mixtures of organic solvents and water have resulted in high quality spectra. Even though these solvents are chemically very different from lipid membranes, it could be shown that the structure in such solvents can be very close to the structure which is adopted under more physiological conditions[355]. A system which is more closely related to biological membranes is the use of small spherical micelles. These micelles are composed of lipid molecules and are therefore chemically very close mimics of biomembranes. Their small size makes them suitable for NMR studies, however, also it results in the main difference with respect to lipid bilayers, which is their high surface curvature. A list of fragments, whose structures have been successfully solved by NMR is compiled in the following table:

Receptor	Fragment	Structure	Function	Medium	Length	Ref.
Angiotensin II ATA1A	Helix 8	amphipathic helix	membrane interaction	water, TFE/water (1:2)	21	[356]
	i3 (res. 213-231)	amphipathic helix	n.d.	TFE/water (1:2)	19	[356]
	i3 (res. 227-242)	mainly disordered	none found	TFE/water (1:2)	16	[356]
	e1 loop	Type 2 β-turn stabilized by membranes	membrane interaction	water, SDS micelles	15	[357,358]
Parathyroid hormone receptor	i3 loop (linear)	N-terminal helix, rest extended	membrane interaction	SDS micelles	29	[359]
	i3 loop (cyclic)	N-terminal helix, rest in loop conformation	membrane interaction	DPC and SDS micelles	29	[359,360]
	e1 loop	Turn with helical stretches on either side of the loop	membrane interaction	DPC micelles	44	[361]
Cannabinoid receptor	i3 loop	Turn with helical stretches on either side of the loop	inhibits interaction with G-protein	SDS micelles	44	[362]
	Helix 8	helical in presence of micelles	membrane interaction	water and DPC micelles	21	[363]
	TM1-i1-TM2	n.d. (assignment only)	n.d.	DMSO	74	[364]
	N-term-TM1-i1-TM2	n.d.	n.d.	DMSO	101	[364]
	TM5-i3	helical TM5	n.d.	DMSO	53	[365]
Cholecystokinin-2 receptor	e3 loop	Turn with helical stretches on either side of the loop	ligand and membrane interaction	DPC micelles	27	[366-368]
α-factor receptor (Ste2p)	TM1	glycine-kinked helix	flexibility in kink region	TFE/water (4:1)	33	[369]
	TM2	bent helix	n.d.	TFE/water (4:1)	35	[369]
	TM3	straight helix	rigid helix	TFE/water (4:1)	35	[369]
	TM4	bent helix	n.d.	TFE/water (4:1)	36	[369]
	TM5	bent helix	rigid helix	TFE/water (4:1)	36	[369]
α-factor receptor (Ste2p)	TM6	proline-kinked helix	flexibility in kink region	TFE/water (4:1)	31	[369]
	TM7	proline-kinked helix	flexibility in kink region	TFE/water (4:1)	30	[369]

Receptor	Fragment	Structure	Function	Medium	Length	Ref.
α-factor receptor (Ste2p)	e1 loop (linear)	flexible with slight helical tendency	n.d.	DMSO	33	[370]
	e1 loop (cyclic)	flexible with slight helical tendency	n.d.	DMSO	33	[370]
	e3-TM7-C-term	helical TM7, some helical tendency in C-term		TFE/water (1:1)	73	[371]
	e3-TM7-C-term	helical TM7, some helical tendency in C-term		CHCl3/MeOH/water (4:4:1)	73	[371]
	e3-TM7-C-term	helical TM7, unstructured e3 and C-term	micelle integration	DPC micelles	73	[14]
	N-tern-TM1-i1-TM2-e1	N-terminus containing amphipathic helix followed by helix(TM1)-loop(i1)-helix(TM2) motif	micelle integration of TMs and micelle interaction of	LPPG micelles	80	[372]
rhodopsin	C-terminus	second half well defined with short antiparallel β-sheet	inhibits interaction with G protein	not mentioned	33	[354]
	N-term	helical with breaks at prolines	n.d.	DMSO	40	[373]
	e1 loop	Turn with helical stretches on either side of the loop	n.d.	DMSO	32	[373]
	e2 loop	Turn with helical stretches on either side of the loop	n.d.	DMSO	34	[373]
	e3 loop	Turn with helical stretches on either side of the loop	n.d.	DMSO	25	[373]
	TM1	Helical conformation	none found	DMSO	16/16	[347]
	TM4	Helical conformation	none found	DMSO	17/16	[347]
	TM5	helical conformation with disordered ends	none found	DMSO	20/18	[347]
	TM6	helix bent at internal Pro	n.d.	DMSO	15	[374]
	i1 loop	well structured β-turn	none found	water	17	[375]
rhodopsin	C-terminus	second half well defined with short antiparallel β-sheet	inhibits interaction with G protein	not mentioned	33	[354]
	i3 loop	turn-helix-turn motif	inhibits interaction with G protein	water	22	[376]
	Helix 8	helical in presence of vesicles (CD)	none found	SUVs, LUVs	12	[377]
CCR5	N-term	helical from res. 7 to 15	interaction with gp120-CD4 complex	water	14	[303]

Receptor	Fragment	Structure	Function	Medium	Length	Ref.
Bradykinin B2 receptor	i2 loop	Turn with helical stretches on either side of the loop	membrane interaction	DPC	36	[378]
	C-term fragment	Short amphipathic helix (helix 8)	membrane interaction	DPC	58	[379]
Neurokinin-1 receptor	N-term fragment	helical structure near beginning of TM1	none found	DPC	39	[380]
	e3 loop	helical in presence of micelles	membrane interaction	DPC	26	[380]
	e2 loop fragment	helical in presence of micelles	none found	DPC	37	[381,382]
β-adrenergic receptor	i3 loop	helical structure near beginning of TM6 in presence of micelles	membrane interaction	TFE, LMPC	12	[383]
	C-term fragment	Helical conformation	unstructured in water, helical in presence of detergents/lipids	water, LMPC, di-MPC	15	[383]
	Helix 8	Helical conformation in DMSO, disordered in water	none found	water, DMSO, DPC	34	[384]
α2A adrenergic receptor	TM3-i2	helical in presence of micelles	membrane interaction	DPC	32	[385]
	TM3-i2-TM4	helical in presence of micelles	membrane interaction	DPC	35	[385]
Adenosine A_{2A} receptor	TM1	Helical conformation (CD)	none found	vesicles and micelles	23	[386]
	TM2	Helical conformation (CD)	none found	vesicles and micelles	23	[386]
	TM3	Helical conformation (CD)	none found	vesicles and micelles	26	[386]
Adenosine A_{2A} receptor	TM4	Helical conformation (CD)	none found	vesicles and micelles	24	[386]
	TM5	Helical conformation (CD)	none found	vesicles and micelles	26	[386]
	TM6	Helical conformation (CD)	none found	vesicles and micelles	27	[386]
	TM7	Helical conformation (CD)	none found	vesicles and micelles	28	[386]
V1A vasopressin receptor	i2 loop	Turn with helical stretches on either side of the loop	inhibits interaction with G protein	TFE, SDS, DPC	22	[375,387]

Receptor	Fragment	Structure	Function	Medium	Length	Ref.
Thromboxane A2 receptor	e2 loop	Loop structure in solution	none found	water	21	[388]
	e3 loop	Loop structure in solution	none found	water	19	[389]
Corticotropin releasing factor receptor	N-terminal domain	Defined 3D structure with partially unstructured regions	Folding of unstructured region upon ligand binding	water	95	[174]

5. The neurohormones of the neuropeptide Y family

5.1 Introduction to the neurohormones of the neuropeptide Y family

Neuropeptide Y (NPY) and the two gut hormones peptide YY (PYY) and pancreatic polypeptide (PP) form the NPY family of neurohormones. Neurohormones are defined as hormones which are produced by or act on the nervous system. All three peptides are 36 amino acids long and carry an amidated carboxy-terminus, which is essential for their functioning (figure 6).

5.2 Pharmacological profiles of the neurohormones from the NPY family

PP was the first family member to be identified and characterized[390]. It is almost exclusively synthesized in endocrine pancreas and released in response to food intake. The known effects of PP are mainly restricted to the gastrointestinal tract and include inhibition of pancreatic secretion and regulation of intestinal motion[391]. Binding sites for PP have, however, also been found in several brain regions of rats[392], suggesting its potential to directly affect the central nervous system. This hypothesis is also supported by the finding that intracerebroventricular injection of PP stimulates feeding in several different species[393,394].

PYY was first isolated from porcine intestine using a method specifically developed for capturing carboxy-terminally amidated peptides[395-398]. Briefly, crude peptide mixtures were treated with a protease and the free amino acids were derivatized with dansyl chloride. Amidated amino acids were then distinguished from free carboxylates based different solubility in organic solvents[395]. Because of the N- and C-terminal residue being tyrosines (Y), the peptide was named peptide YY. Similar to PP also PYY is released in the gastrointestinal tract in response to ingestion and has similar physiological effects[391].

NPY was first isolated from pig brain using the same method by which PYY was discovered[397]. It is one of the strongest orexigenic peptides known. Upon intracerebroventricular injection it has been found to induce carbohydrate-rich food uptake in many species[393,399]. Centrally administered NPY decreases thermogenesis[400], has anticonvulsant activity[401], inhibits sedation[402], mood and

memory[403] and has general anxiolytic effects[404]. NPY is co-localized with noradrenaline in sympathetic nerves and enhances noradrenaline-mediated vasoconstriction[405]. Centrally administered NPY, however, reduces arterial blood pressure and heart tone[406]. NPY is implicated in several pathological conditions: The most obvious are, not surprisingly, based on its orexigenic effect, various eating disorders such as anorexia, bulimia neurosa, and diabetes[407]. Also several cardiovascular disorders and some tumor diseases are associated with abnormal plasma levels of NPY[408]. A very interesting finding indicates that NPY plays a role in alcohol consumption in rats[409,410] and in humans[411], a result particularly interesting in the context of the general anxiolytic activity of NPY. The neurohormone NPY is itself regulated by several neuropeptides and hormones[412]. One example is the 28 amino acid peptide ghrelin that is released in the gut and triggers the release of growth hormone via a GPCR[413,414]. Another example is the adipose hormone leptin which acts as a satiety factor possibly by inhibiting NPY release in the hypothalamus[415]. NPY is the most abundant neuropeptide in the mammalian central nervous system[416], but is also expressed in the peripheral nervous system[417]. Briefly the actions of the three neurohormones of the NPY family can be summarized as NPY having neurotransmitter properties[418] and PYY and PP acting as hormones[419].

5.3 Biosynthetic aspects of the neurohormones of the NPY family

C-terminal amidation is found with roughly 50% of the bioactive peptide hormones. The primary translation products is a precursor containing signals for the appropriate biochemical modification[420]. Amidated peptides are usually synthesized in specifically differentiated secretory cells where the precursors are cleaved into the final products plus a one residue glycine overhang at the C-terminus [420,421]. The formation of the amide is then the last step in the synthesis of the bioactive compounds[420,422]. An amidation enzyme (peptidyl-glycine α-amidating monooxygenase [PAM])[423] and a carboxypeptidase[424] responsible for these two posttranslational modifications of the peptide hormone precursors have been cloned.

5.4 Structural characteristics of the neurohormones of the NPY family

The amino acid sequence of NPY is one of the most conserved during evolution with

22 positions being constant in all known NPY sequences[425]. The number of conserved residues is decreased to 15 in PYY and 7 in PP[425], making the latter one of the least conserved peptides known so far[426]. Seven positions are absolutely conserved among all species of NPY, PYY and PP. These are Pro5, Pro8, Gly9, Ala12, Tyr27, Arg33 and Arg35[425] (figure 6).

The first atomic resolution structure of a member of the NPY family was obtained for avian PP (aPP) by X-ray crystallography[427] and later for bovine PP (bPP) in solution by NMR[428].

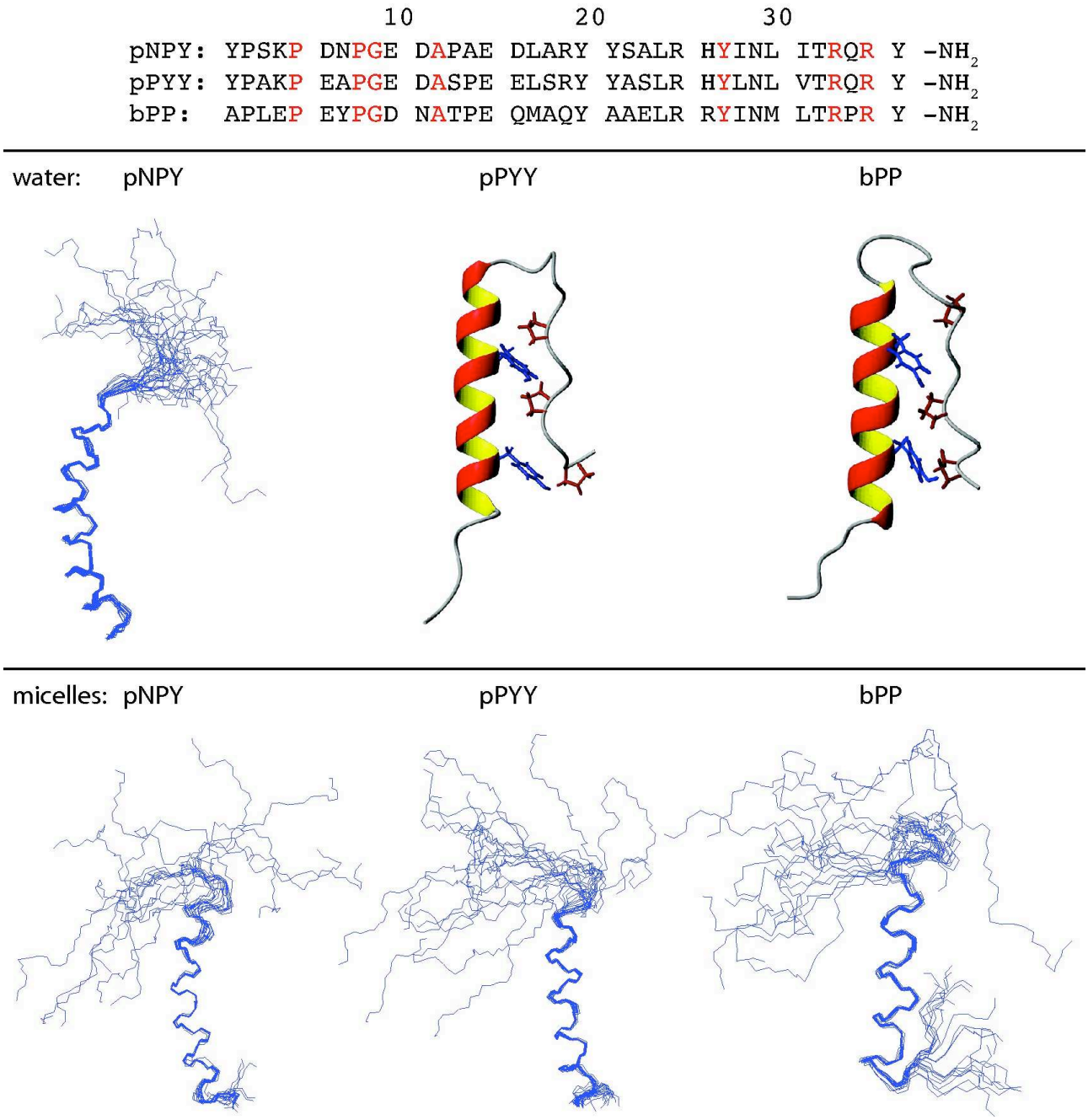


Figure 6: *Neurohormones of the NPY family. Top: Sequence comparison of porcine NPY (pNPY) and PYY (pPYY) with bovine PP (bPP). Residues which are conserved among all three neurohormones throughout all species are colored red. Bottom: Threedimensional structures of pNPY, pPYY and bPP in water and in DPC micelles. In the structure of pPYY and bPP in water, the interdigitating Pro and Tyr residues are colored red and blue, respectively.*

Residues 1 to 8 form a type-II polyproline helix, which is followed by a type-I β-turn connecting to residues 15 to 32, which form an α-helix, and the four most carboxy-terminal residues are in a flexible loop conformation. A surprisingly stable helical hairpin is formed by backfolding of the polyproline helix onto the α-helix. This structural motif is commonly referred to as the PP-fold (figure 6). The tertiary

contacts which stabilize this fold are formed by interdigitation of the conserved proline residues in position 2, 5, and 8 on the polyproline helix and the tyrosines 20 and 27 on the α-helix.

The solution structure of PYY in aqueous solution was shown to be highly similar and to also display the characteristic PP-fold[429,430]. Surprisingly, in the highly homologous NPY the backfolding is absent[431,432] (figure 6). The NMR data which indicated the absence of a stable helical hairpin conformation for human[431] and porcine[432] NPY suggested the presence of NPY dimers which interact via the side chains of their C-terminal α-helices in an anti-parallel fashion. Based on data obtained from CD-spectroscopy it was suggested, that at lower concentrations NPY is present in a backfolded form[433]. According to these data dimer formation is only observed at high concentrations of NPY, like they are typically used for NMR studies, and goes hand in hand with unfolding of the helical hairpin conformation. The assumption of backfolded NPY at low concentrations, however, is controversial, as there are also data arguing against this hypothesis[434]. Dimer formation has also been observed in the crystal structure of aPP[427]. The dissociation constants have been calculated as 1.6 μM porcine NPY (pNPY)[432] and 0.3 μM for aPP[435].

The structures of the neurohormones in aqueous solution in the absence of any biological membranes or cognate receptors might deliver only limited information on their bioactive conformation. It would therefore be desirable to conduct structural studies in the presence of biomembranes, or preferentially, functional receptors. The latter approach requires the ability to produce and purify the receptor in mg amounts and could only be applied to a very limited set of ligand-GPCR pairs (*vide supra*). Structure determination by NMR in presence of biomembranes is usually not feasible, due to the large size of the membrane particles, which are usually present in the form of vesicles, whose slow rotational tumbling is prohibitive for NMR studies. The formation of phospholipid aggregates is dominated by two opposing forces: the electrostatic repulsion between head groups and the attractive van der Waals interactions of the hydrophobic tails. In cases where the repulsive forces are dominating, micelles are formed, as these have the geometry with a minimal number of head groups per surface area. Often such micelles are small enough to allow structure determination of bound molecules by NMR. Micelles resemble biological membranes with respect to their polarity profile[7] and are therefore often referred to as

"membrane mimicking agents". The small size of these spherical particles entails a large surface curvature, which is in contrast to commonly found biological membranes[436]. A large variety of micelle forming (synthetic) detergents are available and have been shown to be suitable for NMR studies[202].

All three members of the NPY family have been characterized in terms of their structure and dynamics in presence of dodecylphosphocholine (DPC) micelles (figure 6). Whereas in aqueous solution PYY and PP are structurally similar, in the presence of micelles PYY[430] and NPY[437] adopt a similar conformation, with PP taking an alternative conformation[438]. It was found that pNPY and pPYY interact via the hydrophobic face of their C-terminal α-helix with the micelle, while the N-terminus freely diffuses in solution. bPP also interacts with the micelle via its C-terminal α-helix, but the N-terminus is also loosely associated with the micelle surface. It has been proposed that this association is mediated by Tyr7 in bPP[430], which has a favorable free energy of transport into the water-membrane interface[439], in contrast to the Asn7 and Ala7 found in pNPY and pPYY, respectively. In addition also the C-terminal pentapeptide, which contains particularly important residues for receptor binding (*vide supra*), differs structurally in bPP from the pNPY/pPYY pair[430].

The NMR study of human parathyroid hormone (hPTH) in the presence of DPC micelles showed that the membrane environment induced a higher degree of conformational order than the one found in purely aqueous buffer[440].

5.5 Ligand recognition from the membrane bound state

In his membrane compartment theory Schwyzer proposed the membrane as a catalyst for peptide-receptor interactions[441-445]. In this theory the membrane is subdivided into three electrostatically distinct compartments: the hydrophobic, the fixed-charge and the aqueous compartment. The membrane influences receptor selection of regulatory peptides by guiding important residues into the appropriate compartment.

The interaction of the neurohormones of NPY family with micelles might imply a model for the binding of the ligand to the receptor, in which in a first step the ligand binds to the membrane, thereby increasing the effective hormone concentration in the vicinity of the ligand and reducing the search for the receptor from three- to two-dimensional space. According to the membrane compartment theory binding to the membrane probably occurs in a conformation facilitating subsequent receptor binding

and hence the membrane bound conformation has been proposed to resemble the receptor bound conformation. In the case of NPY and PYY this hypothesis is corroborated by the similarities of their pharmacological profiles[446] which are reflected by their structural resemblance when bound to micelles, but not in free aqueous solution[430]. In the proposed model for the binding mechanism of NPY family peptides to their receptors the peptides first bind to the membrane through electrostatic interactions. The peptide then reorients such that the hydrophobic residues penetrate into the interior of the membrane. In this state the hormone diffuses along the membrane. In this conformation the peptide can be recognized by the receptor and may or may not undergo further conformational changes to finally adopt the bioactive conformation in the ligand binding pocket of the membrane[430,446].

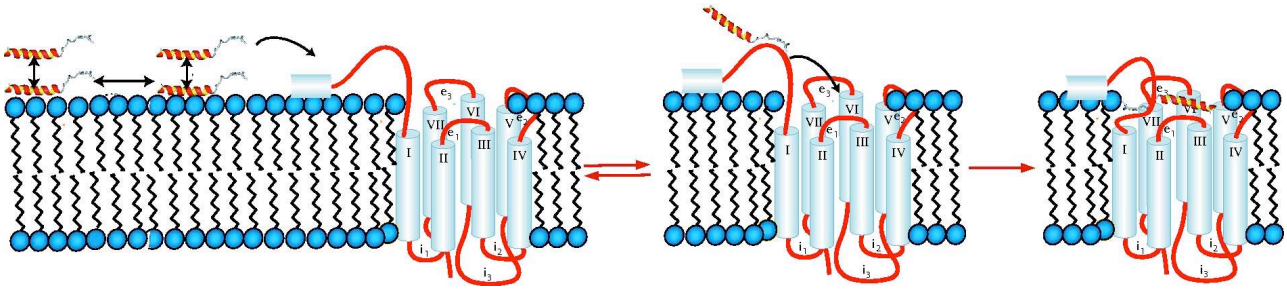


Figure 7: *A model for the recognition of neurohormones from the membrane bound state. An equilibrium between free and membrane bound neurohormone (left) preceeds the formation of preliminary transient contacts with the receptor (middle) and the final insertion of the neurohormone in the receptor's ligand binding pocket.*

5.6 Interaction of the neurohormones of the NPY family with their receptors

In an attempt to identify interacting residues within the ligand Beck-Sickinger *et al.* have conducted a full Ala scan for human NPY at the human Y1-receptor[447]. Most significant reductions in affinity were observed for Arg33 and Arg35. Moreover, binding was almost completely abolished when the C-terminal amide was replaced with a free C terminus. From these results direct interactions involving hormone residues Arg33 and/or Arg35 and an Asp receptor residue (Asp289 in case of the hY1-receptor) have been postulated. Nevertheless, Dougherty has also proposed that pi-cation interactions involving one of the Arg residues and aromatic receptor residues contribute to binding[448].

A turn-inducing dipeptide composed of alanine and amino isobutyric acid (Ala-Aib) introduced into positions 31 and 32 of pNPY, yielding [Ala^{31}, Aib^{32}]-pNPY has been

identified as conferring Y5-receptor selectivity[449]. The same could be observed for the [Ala[31], Pro[32]]-pNPY mutant[450]. The micelle bound structures of these mutants in comparison with wild-type pNPY only differ in the C-terminal residues, which are more disordered in the mutants than in the wild-type.

To understand the different pharmacological profiles and solution structures of NPY and PP, chimeras of these two peptides were studied. Exchanging the segment 19-23 in human PP (hPP) with the sequence found in pNPY, yielding the chimera [pNPY[19-23]]-hPP, had no effect on the formation of the PP-fold as it is also assumed by hPP. The complementary [hPP[19-23]]-pNPY also displayed the structural features of its parent pNPY, which does not assume the PP-fold in solution. When bound to detergent micelles the two chimera showed the same overall structures as did their parent molecules, exhibiting a structured C-terminal helix and a flexible N-terminal tail. The C-terminal part of the helix was identical in the chimeras and the parent peptides, but the N-terminal part differed in the starting points of the helix[451].

In a different study the receptor binding properties of hPP was studied by the introduction of point mutations, the exchange of fragments 1-7, 1-17 and the above mentioned 19-23. A combination of several changes resulted in [cPP[1-7], pNPY[19-23], H[34]]-hPP which showed specificity for the Y5-receptor[452].

6. The Y-receptor subfamily of GPCRs

6.1 Introduction to the Y-receptors

The neuropeptides exert their effects via several receptor subtypes (Y-receptors). Four main receptors, named Y1-[453-455], Y2-[456-459], Y4-[460,461] and Y5-[462], have been cloned so far. Two minor receptors named Y3-[463] and y6-[464] receptors have been cloned, but little information is available for them. In the following I will mainly focus on the four main Y-receptors. All identified Y-receptors belong to the superfamily of G protein-coupled receptors (GPCRs) and act via G proteins of the G_i family, whose activation leads to an inhibition of the adenylate cyclase. They range in size from 375 to 455 residues and show the prototypical characteristics of GPCRs: They show seven hydrophobic amino acid stretches, which are spanning the membrane as α-helices.

These transmembrane helices are connected by three intra- and three extracellular loops. Typical for the subfamily 1b of GPCRs there is a predicted disulfide bond between extracellular loops 1 and 2. Furthermore soluble, extracellular N-terminal domain and intracellular C-terminal domains are present.

6.2 Pharmacological profiles of the Y-receptors

The different receptor subtypes are localized in various tissues, both in the central nervous system and in the periphery.

The table below gives a summary about the most important data concerning the four major Y-receptors:

Receptor	**Y1**	**Y2**	**Y4**	**Y5**
K_i NPY [nM]	0.81[a]	0.02[a]	1.9[a]	0.19[a]
K_i PYY [nM]	1.1[a]	0.01[a]	1.1[a]	0.06[a]
K_i PP [nM]	>100[a, b]	>1000[a, b]	0.04[a, b]	27[a, b]
amino acids	384	381	375	455
major occurrence	periphery[c], hypothalamus[d]	CNS[e, f, g]	intestine, pancreas[h]	hypothalamus[i]
related action	vasoconstriction, anxiolysis[c, d, j]	memory, epilepsy, secretionc[k, l]	gastro-intestinal regulation[m]	food intake[i]

Table 1: *Biochemical, biophysical and physiological characteristics of the four major families of the Y-receptors. a) McCrea, Regul. Pept., **87**, 47-58, 2000; b) Small, Proc. Natl. Acad. Sci. U S A, **94**, 11686-91, 1997; c) Wahlestedt, Med. Biol., **64**, 85-8, 1986; d) Wahlestedt, Science, **259**, 528-31, 1993; e) Gehlert, Mol. Pharmacol., **49**, 224-8, 1996; f) Gerald, J. Biol. Chem., **270**, 26758-61, 1995; g) Rose, J. Biol. Chem., **270**, 22661-4, 1995; h) Lundell, J. Biol. Chem., **270**, 29123-8, 1995; i) Gerald, Nature, **382**, 168-71, 1996; j) Grundemar, Br. J. Pharmacol., **105**, 45-50, 1992; k) Flood, Peptides, **10**, 963-6, 1989; l) Potter, Regul. Pept., **25**, 167-77, 1989; m) Schwartz, Gastroenterology, **85**, 1411-25, 1983*

6.3 Evolutionary relationships between the Y-receptors

Sequence comparisons show, that among these four receptors, the Y1 and Y4 are the most closely related (42% sequence homology) while Y2 and Y5 are equally distant from each other and from the Y1/Y4 pair (~30% homology to Y1)[465,466]. These relationships are also reflected when the predicted soluble N-termini of the four receptors are compared, as is shown in the sequence alignments below.

```
NPY1R_HUMAN      MN-STLFSQVENHSVHS-NFSEKNAQLLAFENDDCHLPLAMI 40
NPY4R_HUMAN      MNTSHLLALLLPKSPQGENRSKPLGTPYNFS-EHCQDSVDVM 41
                 ** * *:: :  :* :. * *:  .    *. :.*: .: ::

NPY2R_HUMAN      MGPIG-AEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELID-STKLIEVQ
50
NPY5R_HUMAN      MSFYSKQDYNMDLELDEYYNKTLATENNTAATRNSD-FPVWDDYKSSVDDLQ
51
                 *.  .  : : :  ::*   :  ..:... .   .*  *   * .:.: ::*
```

In several studies the N-termini of Y-receptors have – as is typical for GPCRs of the subfamily 1b – been shown to be involved in ligand binding[467,468]. We therefore sought to express the 40 to 50 residue N-terminal fragments of the four Y-receptors and characterize them NMR-spectroscopically.

6.4 Ligand binding specificities of the Y-receptors

NPY and PYY bind equally well to the receptors Y1, Y2, and Y5 (nanomolar to subnamomolar dissociation constants). Only PP shows selectivity towards the Y4-receptor (picomolar dissociation constant). The ability of NPY and PYY to bind to three different receptor subtypes is probably related to their conformational flexibility, which enables the peptides to adopt more than one energetically favorable conformation.

The Y1-receptor is expressed in blood vessel and parts of the central nervous system. Its most important effects are vasoconstriction[469,470] and anxiolysis[471]. Roles in the regulation of feeding behavior[394] and alcohol consumption[409,410] have been reported. The Y1-receptor is pharmacologically characterized by high affinity for NPY and PYY and the corresponding analogs carrying a proline at position 34 (Pro34 analogs) and low affinity for N-terminally truncated analogs and for PP[453,454,472].

The Y2-receptor is expressed in various locations of the peripheral nervous system, such as sympathetic and parasympathetic nerve fibers, the intestine and certain blood vessels. Its physiological functions are suppression of neurotransmitter release[470,472] and enhancing memory retention[473]. NPY, PYY and their C-terminal fragments are potent activators of the Y2-receptor, while the Pro34 analogs and PP show only weak effects[458,459,462].

The Y3-receptor has been mainly localized in the brainstem. It is involved in

inhibition of catecholamine release[474] and the modulation of arterial blood pressure[475,476]. The Y3-receptor is able to bind NPY and its Pro34 analog, but is insensitive to PYY and PP[475,476].

The Y4-receptor is expressed in various tissues such as the heart, intestine, colon and pancreas. Upon activation it leads to a decrease in pancreatic secretion and contraction of the gall bladder[477]. PP binds with much higher affinity (picomolar dissociation constant) to the Y4-receptor than NPY, PYY and their corresponding Pro34 analogs (nanomolar dissociation constants)[460,478].

The Y5-receptor is expressed in the hypothalamus, where it is believed to play a role in the regulation of food intake[462]. Its ligands include NPY, PYY, their Pro34 analogs and the large N-terminally truncated analogs like NPY(2-36) and NPY(3-36). The C-terminal fragments and PP show reduced affinity[462,479].

The role of the y6-receptor is unclear, because it is a pseudogene in several species[480]. The pharmacological effects of this receptor are controversial[464,481].

The ability of the neurohormones to bind to several different receptors complicates the determination of structure-affinity and structure-activity relationships. For being able to characterize one receptor with respect to the others it would be helpful to have selective ligands. The non-selectivity of receptor binding of NPY and PYY is believed to be a result of their conformational flexibility. Accordingly the ligands are able to fulfill different conformational requirements imposed by the different receptors. Therefore imposing conformational restraints on the ligands might be a potential route to selective ligands for the receptor subtypes. If one were able to synthesize an analog mimicking the active conformation of a ligand on a given receptor, it should be selective for that particular receptor subtype (i.e. show no affinity for the other receptor subtypes). Such analogs can be obtained in the form of non-peptide species or peptides with reduced flexibility, e.g. through cyclization or using spacer templates[482]. Unfortunately efforts to towards selective agonists for GPCRs has been less successful than to obtaining selective antagonists[483], probably because the non-peptidic analogs that were studied were too small and therefore could not induce a conformational change of the receptor, due to a lack of contact points between ligand and receptor.

The molecular evolution of the Y-receptors has been difficult to deduce, due to the lack of available non-mammalian sequences. Presently it is believed that Y-receptors can be sorted into three subfamilies based on their sequence homology. The

subfamilies are named after their first member as Y1, Y2 and Y5. Their overall sequence identity is low (27-31%) and a bit higher in the TM regions (40-43%). The Y1 subfamily also includes the Y4R and y6R and share a sequence identity of 50% and 60% overall and in the TM regions, respectively. Structural comparison suggests that the Y1R/Y4R pair and the Y2R/Y5R pair form distinct subfamilies within the Y-receptor family[484].

All currently known Y-receptors show some absolutely conserved residues. Particularly interesting in terms of ligand binding might be those conserved sites in the extracellular domain, and among those the ones which are Y-receptor specific (i.e. which are only conserved in Y-receptors, but not in all other peptide ligand binding GPCRs) should be determining the binding modes of the NPY family peptides to their receptors.

The search of such residues resulted in the identification of $Asp^{6.59}$, located in the interface between TM6 and ECL3[485]. Interestingly it was found that in other peptide binding GPCRs, position 6.59, together with the adjacent position 6.58, is frequently conserved within a family. The identification of $Asp^{6.59}$ as a residue which might be important in determining the ligand binding mode of Y-receptors was in agreement with previous studies confirming the important role of that residue for ligand binding[467,486,487]. The replacement of $Asp^{6.59}$ with Glu was tolerated in all Y-receptor subtypes, whereas replacement with Ala, Asn and Arg led to a loss in affinity and potency[485]. In a complementary mutagenesis approach the residues of NPY interacting with $Asp^{6.59}$ were identified: Whereas in Y2R and Y5R $Asp^{6.59}$ interacts with Arg^{33} of NPY, in Y1R and Y4R $Asp^{6.59}$ interacts with Arg^{35}.

6.4.1 The Y1-receptor

One of the most thorough ways of characterizing the interaction of a peptide ligand with its receptor is the systematic single exchange of each residue of the ligand by L-alanine. Such a study was carried out to characterize the binding of NPY to the Y1-receptor[447]. The four natural alanines were substituted by glycines. The most important residues for binding of NPY to the Y1-receptor were according to this study Pro5, Pro8, Arg19, Tyr20 and the C-terminal positions 27-36 including Tyr27, Arg33 and Arg35. Whereas for the two C-terminal Arg residues the reduction in binding affinity was on the order of 10^4 to 10^5-fold, for the other mentioned residues it was

around 10^3-fold.

The finding that three positively charged arginine residues are critical for binding of NPY to the Y1-receptor led to the theory that the receptor-ligand interaction might be predominantly electrostatic and therefore mediated through negatively charged residues on the receptor. This hypothesis was tested by replacing negatively charged residues in the extracellular domain of the Y1-receptor with alanines[486]. These mutants were expressed transiently in HeLa cells (epithelial cells of a cervical carcinoma of a patient called **He**nrietta **La**cks) and their ability to bind NPY was assayed by a radioligand binding assay using ^{125}I-NPY[488]. The human Y1-receptor (hY1) contains 30 negatively charged residues of which 14 are located in a putative extracellular domain: 5 in the N-terminal part, 2 in extracellular loop 1 (ECL1), 6 in ECL2 and 1 in ECL3.

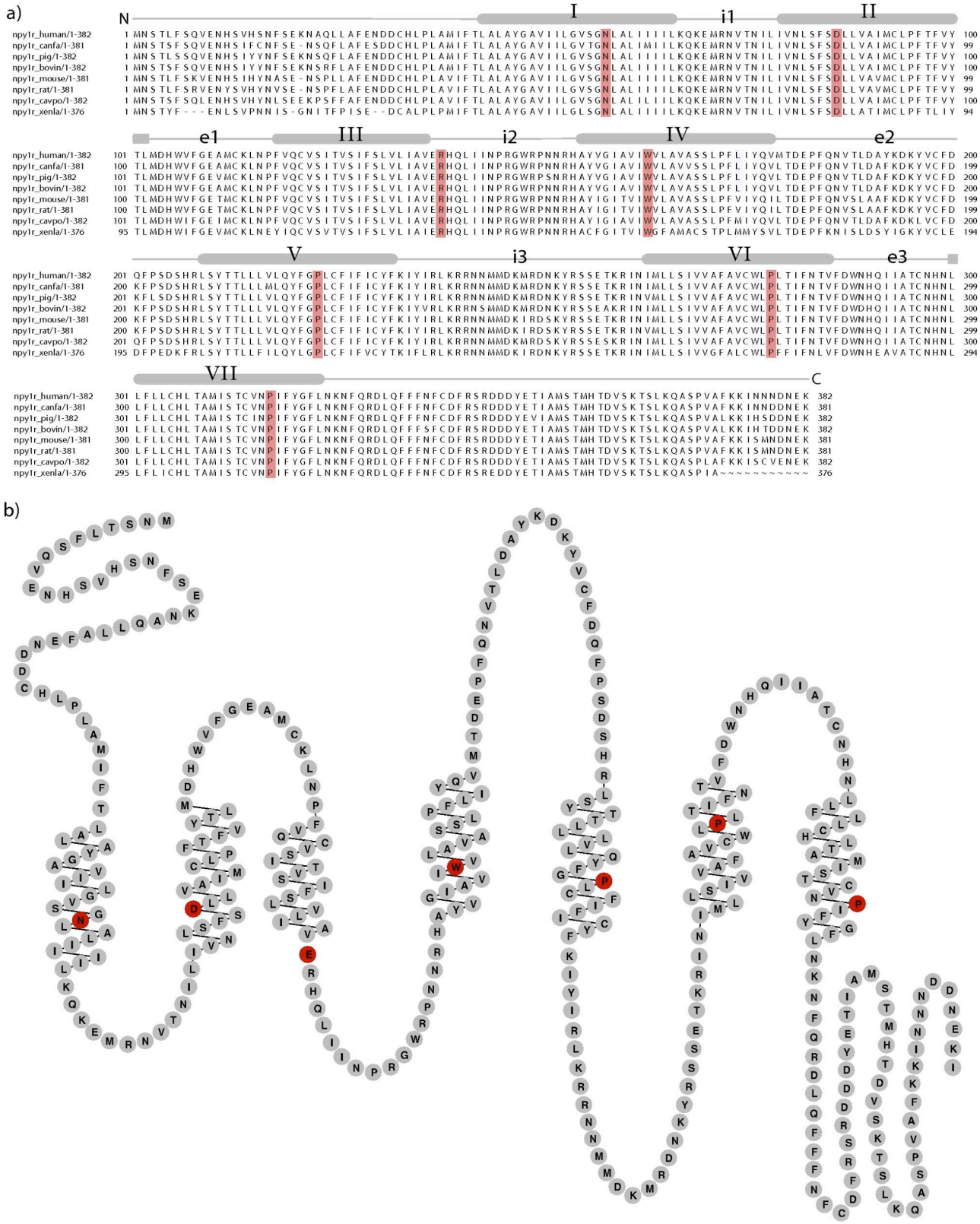


Figure 8: *a) Multiple sequence alignment of the Y1-receptor from eight major mammalian species. Above the sequence the predicted TM helices are indicated by gray rods and all other domains by gray lines. The most conserved residue in each helix (shaded red) serves as the basis for the Ballesteros-Weinstein numbering scheme. b) Snake plot of the predicted topology of the human Y1-receptor. The most conserved residue in each helix (shaded red) serves as the basis for the Ballesteros-Weinstein numbering scheme.*

None of the mutations introduced in the N-terminus affected NPY binding. In ECL1 the D104A substitution led to a complete loss of affinity for NPY, whereas the E110A mutant didn't show any altered NPY binding. In ECL2 two aspartates, namely Asp194 and Asp200, were essential for NPY binding and the mutation of Asp205 resulted in a significant loss of binding affinity. The other 3 positions proved to be tolerant for substitution. The D287A mutation in the ECL3 again resulted in the complete loss of affinity for NPY. It is interesting to note that the mutations affecting NPY binding the most are all clustered at the top of TM helices 5 and 6.

6.4.2 The Y2-receptor

The same Ala-scan which was conducted for the interaction of NPY with the Y1-receptor showed only minor effects on the binding to the Y2-receptor[447]. In the N-terminal part only the replacement of Pro5 lead to a significant (~500-fold) decrease in receptor affinity. The replacement of Arg33 and Arg35 lead to a decrease in affinity of ~10^3- and 10^5-fold, respectively. Surprisingly also the substitution of Thr32 lead to a drop in the affinity of 10^3-fold.

A study in which each residue in NPY was exchanged with its D-isomer showed only reduced affinity for species carrying the mutation in the 30-35 region[489]. Overall the Ala- and D-amino scan showed that the most important part of NPY for binding to the Y2-receptor is the C-terminal fragments. The minimum binding motif necessary for interaction was found to be the C-terminal nonapeptide of NPY[490]. In order to force a defined structure upon this peptide lactam bridges connecting residues i and i+4 were introduced at different positions. It was shown that peptides carrying the bridge between residues 27 and 31 or 28 and 32, respectively, are very potent activators of the Y2-receptors[491]. These cyclic peptides were both studied by NMR and molecular dynamics simulations[491]. Based on these findings it was suggested that the bioactive conformation of the NPY ligand when bound to the Y2-receptor may consist of a hairpin-like structure in which the N- and C-termini are in close proximity, with the C-terminal part being in direct contact with the receptor, and the N-terminal part helping in the stabilization and correct orientation of the C-terminus.

6.4.3 The Y4-receptor

The characteristic feature of the Y4-receptor is its high affinity and specificity for PP showing a K_d in the picomolar range and a more than 100-fold preference over NPY and PP[478]. However, it should be noted that also NPY and PYY display comparably high affinity for the Y4-receptor with dissociation constants in the nanomolar range[492]. The key residues conferring high affinity binding of NPY to the Y4-receptor were determined in a complete Ala scan of NPY[493]. As for the Y1- and Y2-receptors the two Arg residues at positions 33 and 35 where found to be critical for binding. Ala substitution at these sites led to a K_d >1000 nM. Also the replacement of Tyr27 led to an almost 100-fold decrease in the affinity. Binding studies with PP/NPY chimera indicated that the ligand requires a stable helix for being able to bind to the receptor[452].

6.4.4 The Y5-receptor

A lot of research has been dedicated to elucidating the structure-affinity and structure-activity relationships of the Y5-receptor after reports indicated its role in the regulation of feeding behavior, together with the Y1-receptor[494-498].

The Ala scan of NPY revealed that the N-terminal prolines 2, 5 and 8 are important for receptor binding. Their Ala substitution resulted in up to 100-fold losses in binding affinity[493]. The two C-terminal Arg 33 and 35 and Tyr27 were all shown to be important for binding to the Y5-receptor as well. Studies with PP/NPY chimera indicated the requirement of a stable α-helix for receptor binding, which, however, was not sufficient for conferring high affinity[452]. The introduction of an Ala-Aib (Amino isobutyric acid) at positions 31 and 32 of NPY led to a highly selective ligand for the Y5-receptor[449]. It has been speculated that this motif may be responsible for selectivity by inducing a well-defined conformation in the C-terminus of NPY.

6.5 Summary

In summary it can be said that Y1R, Y2R and Y5R preferentially bind NPY and PYY, whereas Y4R shows preference for PP[456,460,461,466,499]. While Y1R and Y4R only can bind full-length ligands[460,500,501], Y5R tolerates the deletion of the first residue[462] and Y2R is even capable of binding significantly shortened peptides such as the NPY(13-

36) variant[500]. The C-terminal pentapeptide of the ligands has been identified to be essential for binding to all Y-receptor subtypes[447,502]. The binding mode of the neuropeptide hormones is different in the Y1/Y4 pair than in the Y2/Y5[485], because of the different orientation of the ligands towards the essential Asp[6.59].

7. Introduction to the loop grafting strategy

7.1 General

Our group has been studying a wide variety of neurohormones of the NPY family using high-resolution NMR[14,430,437,438,446,450,451] and has – based on these studies – proposed a model for the binding of these hormones to their cognate receptors[446]. Structural and mechanistic insight into such binding processes are anticipated to provide a basis both for the rational design of novel and the improvement of already existing agonists and antagonists to a rapidly growing pool of GPCRs for which natural ligands have been identified. Because of the detailed studies that have been conducted in our group on the above mentioned neurohormones and the detailed biophysical and biochemical characterization in other groups, the receptors for the NPY-family neuropeptides are an interesting model system for the investigation of ligand binding to GPCRs for our group.

In many photoaffinity-labeling studies the contact points between receptors and their ligands have been assigned to the extracellular face of GPCRs, which is comprised of an N-terminal domain and three extracellular loops (ECLs) of variable size[381,503,504]. These findings are in accordance with an established model for ligand binding to their GPCRs, which is proposed to consist of an initial membrane-binding event prior to recognition and binding of the receptor[441,446,505]. Structural and dynamical features of the extracellular domains of the receptor are therefore inferred to be of critical importance in the ligand binding process. Even though rhodopsin has served as a starting point for the modeling of many other GPCRs[506,507], doubts about the accuracy of this approach can be raised[508,509]. Especially for the extracellular loops two major concerns apply: first, the loop regions of rhodopsin itself are structurally not well defined in the crystal structure[131] and second, the high structural and functional

diversity of GPCR ligands points to high structural diversity in the receptor loop regions, which is also reflected by a low sequence homology between different GPCRs outside the transmembrane regions.

Even though the production of GPCRs amenable to biophysical studies has made considerable progress in the past[48], the elucidation of membrane protein structures is still a far from trivial task. Their lipidic surrounding prevents crystal growth for x-ray diffraction studies in many cases and increases their effective molecular weight to a range outside that commonly applicable to NMR studies. Alternative approaches to gaining structural information on such difficult targets would therefore be of great importance.

One such strategy, called 'segmentation approach', is the expression of several peptides spanning the whole primary sequence of a target GPCR. It emerged after the finding that small regions of some proteins exhibit the same secondary structure in individual solubilized peptides as observed in the corresponding region of the intact protein[510-512]. Several GPCRs have thus far been characterized in this way[349,513].

We propose here an additional such 'alternative' approach to structural information on the extracellular part of GPCRs. The central idea of our approach is the fusion of the extracellular loops and N-terminal domain of a GPCR onto a soluble protein scaffold, and it will therefore be referred to as 'grafting approach' in the following (see figure 1). Several advantages as compared to the 'segmentation approach' are noteworthy: Binding affinities of ligands for isolated receptor loops often lie in the μM range[349,513]. The binding affinities between ligands and whole receptors, however, commonly are considerably higher, mostly in the nM range (see http://www.gpcr.org/7tm for a collection of ligand binding data). These higher binding affinities can be explained by the cumulative effects of two or more μM binding sites present in two separate loops. A system in which several or all loops of one receptor can be presented simultaneously would therefore seem desirable; a feature provided by the 'grafting approach', at least under the assumption that a scaffold with the desired geometry can be found. An additional requirement for a suitable scaffold is its solubility, which helps in studying the often partly hydrophobic loops.

8. The scaffolds

8.1 Lipocalins

8.1.1 Nomenclature, family members and classification of the lipocalins

The lipocalins have first been identified in a study comparing the sequence of the milk protein β-lactoglobulin (BLG) of two aquatic mammals (dolphin and manatee) with bovine BLG. In this study the similarity to human serum retinol binding protein (RBP) was noted, and conserved sequence and structural motifs were identified[514]. In the subsequent years numerous eukaryotic proteins were assigned to the newly discovered protein family. It was only several years later that members of the lipocalin family were also identified in prokaryotes[515,516].

For the classification of lipocalins two systems have been proposed. One relies on the presence or absence of conserved sequence motifs and divides the members accordingly into kernel and outlier lipocalins[517]. Kernel lipocalins, which through this classification system make up the largest subgroup, each share three conserved sequence motifs, while members of the family matching not more than two of these motifs are falling into the outlier group. The first of these motifs is shared by all lipocalins and can therefore also serve as a means of assigning a protein to the lipocalin family.

However, due to the unusually low sequence similarity among different lipocalin-family members a classification, which is not based on sequence information seems desirable.

Therefore Skerra has proposed a classification based on the knowledge of the three dimensional structure of several lipocalin family members. A group of six biochemically different proteins, coinciding very much in their three dimensional structure are named the “prototypic” lipocalins[518]. Proteins having a lower root mean square deviation (RMSD) than a given threshold to the above mentioned group of sequences will accordingly also fall into the class of prototypic lipocalins.

8.2 Sequence and structure relationships of the lipocalins

The lipocalin fold is very well conserved and characterized by an eight-stranded β–barrel and a C-terminal α-helix[519]. The eight β-strands of the barrel (commonly referred to as strands A to H) are linked by a succession of +1 connections. These seven loops (commonly referred to L1 to L7) are all short β-hairpins, except loop L1 which is a large Ω-loop folding back to close partially the ligand binding-site found at its side of the barrel. Preceding strand A the lipocalins carry a characteristic 3_{10}-helix (figure 10a and 10b).

With the exception of bovine odorant binding protein all mammalian lipocalins carry between one and three disulfide bridges.

The lipocalins form together with two other families of ligand-binding proteins, namely the fatty acid binding proteins (FABPs) and the avidins, the superfamily of calycin proteins[520]. The avidins are basically very similar to the lipocalins, that is they are also eight-stranded β-barrels. The differences between the two families are the lacking C-terminal α-helix and the replacement of the Ω-L1 through a normal β-hairpin in the avidins. The FABPs are structurally somewhat different. They are ten-stranded β-barrels also lacking the C-terminal α-helix and carrying a long L1 loop with two short α-helices in it.

The primary sequence, however, shows unusually low levels of sequence conservation among different members of the family. Only few characteristic signatures could be detected. The most general being a GXW motif near the N-terminus[519,521].

8.3 Biological functions of lipocalins

Lipocalins are typically small (15-20 kDa) secreted proteins which often show the ability to bind small hydrophobic molecules (e.g. retinol, as mentioned above).

The biological functions of the lipocalins goes far beyond transport functions initially assigned to them. Until now such diverse tasks as pheromone activity[522] cryptic coloration[523], olfaction[524], enzymatic synthesis of prostaglandins[525,526], and even cellular regulative functions such as regulation of the immune response[527,528] and mediation of cell homoeostasis[529], have been identified as being accomplished or

regulated by lipocalins (for a review of lipocalin functions see Flower, 1996[517].

8.4 Bacterial lipocalins

The first bacterial lipocalin Blc was identified in *Escherichia coli*[515]. Blc was identified as contributing for the adaptation of cells to starvation and a high-osmolarity environment. Both of these conditions are known to exert particular stress on the cell envelope.

Blc is translated with a type-2 signal peptide (also referred to as lipoprotein signal peptide) that is characteristic for bacterial lipoproteins[530]. This signal peptide is made up of a N-terminal sequence, which carries formyl-methionine and it contains one or several basic amino acids (N-region) followed by a hydrophobic segment (H-region). These two regions are common in all types of signal peptides. Lipoprotein signal peptides are followed by a conserved cleavage site of three amino acids preceding an invariant cysteine demarcating the mature N-terminus of the protein. This N-terminal cysteine is modified with a N-acyl-S-sn-1,2-diacylglycerylcysteine moiety in *E. coli* mature lipoproteins. Blc lipoprotein is targeted to the outer membrane and thought to be exposed to the periplasm[531].

Because of the lack of the second structurally conserved region that is found in all 'kernel' lipocalins, Blc has been classified as an 'outlier' lipocalin[516].

From a synthetic standpoint it is important to note, that Blc is also different from most lipocalins in that it lacks intramolecular disulfide bonds facilitating its handling.

Blc shows close similarity to the mammalian lipocalin apolipoprotein D[515] commonly referred to as ApoD and to the insect lipocalin lazarillo of the American grasshopper *Schistocerca americana*[532] (figure 9).

Blc, ApoD and lazarillo are the only lipocalins known so far to be anchored to lipid bilayers. While in Blc and in lazarillo this task is accomplished by the fatty-acylated cysteine and a C-terminal glycosylphosphatidylinositol (GPI) anchor respectively, ApoD is thought to bind to membranes via its hydrophobic loop inserted between β-strands 7 and 8[533]. The sequence alignment of ApoD with Blc shows, that this hydrophobic loop is the only significant insertion of amino acids (see figure 9). ApoD is thought to interact with high-density lipoprotein (HDL) particles in blood plasma

by means of its hydrophobic loop inserted in between β-strands 7 and 8[533]. Supporting this model is a lone cysteine residue found adjacent to this putative hydrophobic binding loop which was found to form an intermolecular disulfide bond with the HDL-associated apolipoprotein A2[534].

The alignment of lazarillo with Blc and ApoD (see figure 9) shows that the amino acids corresponding to the putative hydrophobic binding loop in ApoD are mutated in lazarillo to hydrophilic residues, which might reflect its different ligand binding requirements.

The similarity between the three proteins mentioned above is, however, not restricted to structural features, but also a close relationship on the functional level can be observed:

While lazarillo is mainly expressed in developing neurons and is thought to control axon guidance[535], ApoD accumulated in remyelinating peripheral nerve following damage[536] and Blc is, as mentioned before, expressed upon stress on the cell envelope. This suggests a common biological role of these three proteins in the repair, remodeling and biogenesis of membranes.

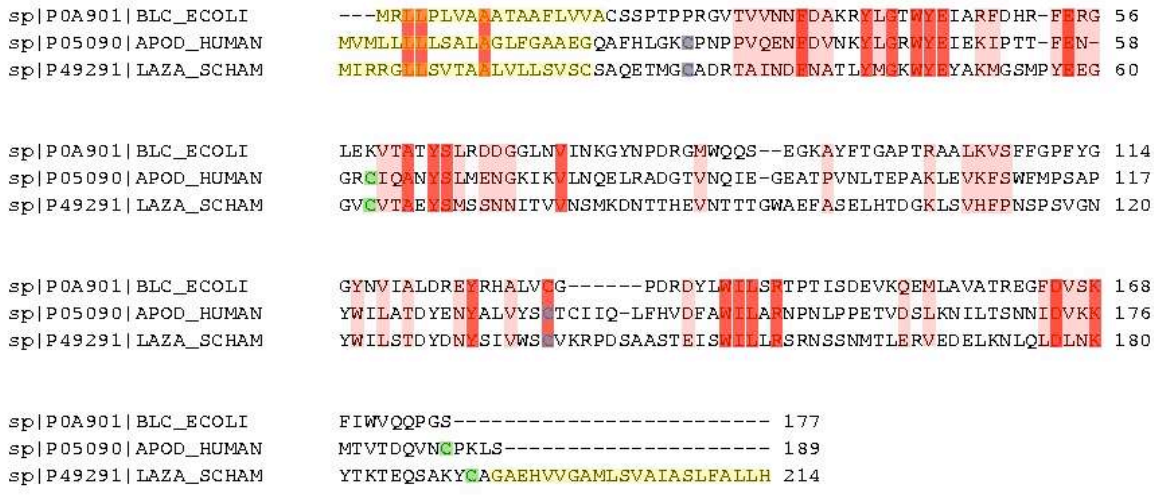


Figure 9: *Sequence alignment of bacterial lipocalin, human apolipoprotein D and the lazarillo protein. Absolutely conserved and conserved residues are indicated by solid and faint red shading, respectively. Sequences which are cleaved off in the mature proteins are shaded yellow and disulfide forming cysteines are shaded pairwise in purple (disulfide bond 1) and green (disulfide bond 2). The hydrophobic stretch anchoring Apo D to the membrane extends from C136 to L140.*

8.5 Lipocalins as Scaffolds

The term scaffold designates a (natural) molecular architecture onto which structurally unrelated elements can be grafted with the goal of creating novel functions. Protein scaffolds are characterized by a stable core region, which is capable of folding irrespective the peptide sequence outside this region. In an ideal case the region outside the core is completely uncoupled from the folding of the stable core and tolerates every conceivable sequence change. A few types of protein folds have been identified and evaluated as potential scaffold candidates[537]. In nature this type of modular protein architecture is rare. Prominent examples are the immunoglobulins (Igs)[538], the TIM barrel enzymes[539], and in some respect also the GPCRs (*vide supra*). The identification of new protein scaffolds is driven by the search for artificial proteins with tunable, prescribed ligand binding properties. The antibodies of our adaptive immune system constitute such library of binding molecules, made up six hypervariable loops displayed on a conserved structural framework. An estimated 10^8 antibodies are circulating through the human body allowing the tagging of intruding structures for degradation/clearance by dedicated host cells.

Retinol-binding protein (RBP) was the first lipocalin, whose high resolution structure was solved[540,541]. Soon after the structure of the lipocalin bilin-binding protein (BBP) was elucidated[542,543]. It was obvious that despite the low sequence homology, structurally these two proteins where closely related. The structures of four additional lipocalins corroborated this view[544-547]. A superposition of 5 lipocalins is shown in figure 10c.

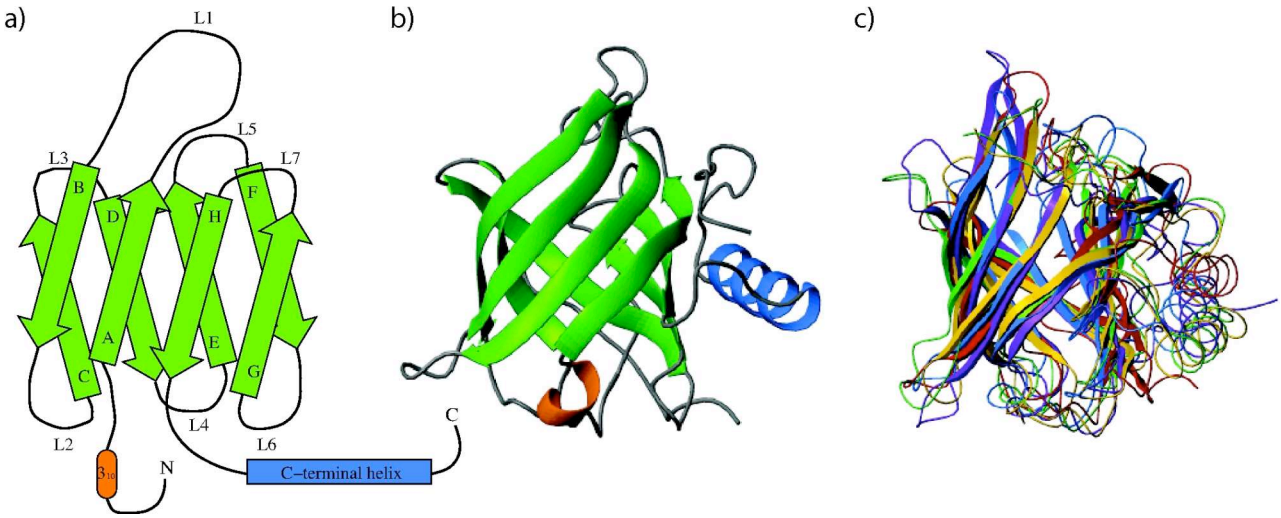


Figure 10: *a) The characteristic structural features of a lipocalin are eight β-strands (green) forming a barrel, an N-terminal 3_{10}-helix (red) and a C-terminal α-helix (blue). b) The crystal structure of the bacterial lipocalin (pdb code 1QWD) with the characteristic structural elements colored as described under a). c) Superposition of 5 lipocalins (bilin binding protein [1BBP]: yellow; bacterial lipocalin [1QWD]: red; retinol binding protein [1RBP]: purple; bovine β-lactoglobulin [1B8E]: blue; apolipoprotein D [2APD]: green). The β-barrel core is structurally conserved, while the loops show large structural variations.*

Based on high structural homology within the β-barrel core and considerable structural and sequence divergence within the loop regions, lipocalins have been proposed as molecular scaffolds. In a first application of this idea a metal-binding site was introduced on the solvent-exposed surface of RBP[548]. A few years later also the loops at the open end of the barrel were changed by a combination of rational protein design and combinatorial protein biochemistry to yield a protein capable of binding fluorescein[549]. Several positions in the BBP sequence were randomly mutated to generate a large library of different sequences. Individual mutants displaying fluorescein binding properties were selected from this pool using phage display technology[550]. In an analogous screening, lipocalin-derived binders to the biologically important molecule digoxigenin could be generated[551]. Lipocalins have, however, also been successfully used as a scaffold for generating high affinity binders to human proteins, as has been shown for an engineered apolipoprotein D binding to hemoglobin[552]. Such binders were termed "anticalins", because they mimic the antibodies of the immune system.

The physiological role of lipocalins is the binding and sequestration/transport of hydrophobic and/or chemically sensitive molecules[517,537]. In this respect the lipocalins seem to be ideal candidates for creating artificial binders to small molecules, an

activity which has proven difficult with many other scaffolds.
The availability of a high resolution structure is a prerequisite for the use of any given protein as a scaffold. Additional practical aspects, such as the presence/absence of posttranslational modifications or disulfide bonds, should also be considered, when selecting a scaffold.

9. Outer membrane protein A of *E. Coli*

9.1 The outer membrane of gram negative bacteria

Bacteria are commonly classified into two or three categories according to their appearance after treatment with the so-called Gram staining procedure[553]. Gram-positive bacteria get stained by the Gram staining procedure. They have a thick peptidoglycan layer (*vide infra*) surrounding their plasma membrane. Gram-negative bacteria, on the other hand, are not colored by the applied stain. Their plasma membrane is also enveloped by a peptidoglycan layer, which itself, however is surrounded by an additional lipid bilayer. Gram-negative bacteria are therefore characterized by two lipid bilayers, one called the inner membrane (IM) and one called the outer membrane (OM). The space between these two membranes is called the periplasm and contains peptidoglycan layer. The peptidoglycan layer is composed of peptide linked polysaccharides: The polysaccharide is a polymer of a disaccharide composed of N-acetyl-muraminic acid and N-acetyl-glucosamine as the repeating unit. Polysaccharide strands are interconnected between N-acetyl-muraminic acid moieties with peptides built up of D- and L-amino acids[554]. The peptidoglycan layer serves as an extracytoplasmic cytoskeleton which determines the cell's shape and protects it from lysis in environments of low osmolarity.

9.1.1 Chemical composition of the bacterial outer membrane

While the IM is a regular phospholipid bilayer (in *E. coli* composed of 70-80% phosphatidylethanolamine, 15-20% phosphatidylglycerol and <5% cardiolipin)[555], the OM is a highly asymmetric bilayer consisting of phospholipids (70-80%

phosphatidylethanolamine, 20-30% phosphatidylglycerol) on the periplasmic and of lipopolysaccharides (LPS) on the extracellular surface[556,557].

LPS are large fatty-acylated oligosaccharides which can be subdivided into three structural parts: The highly immunogenic lipid A, the core oligosaccharide, and the O-antigen polysaccharide[558]. While the core oligosaccharide and the O-antigen are pure saccharide structures, the lipid A molecule is a β-1',6-linked glucosamine disaccharide, which is monophosphorylated on both glucosamine moieties (at positions 1 and 4'). Furthermore it is acylated with four β-hydroxyacyl groups, two or three of which are themselves acylated at their β-hydroxy group with acyl chains[559].

This structural difference is also reflected in the proteins residing in the two membranes: While integral inner membrane proteins have α-helical TM anchors (White, J Biol Chem, 276, 32395, 2001), the outer membrane proteins (OMPs) have β-barrel TM anchors consisting of antiparallel, amphipathic β-strands[560]. Apart from the integral membrane proteins, the outer IM leaflet and the inner OM leaflet also contain lipoproteins, which are anchored to the membrane by an N-terminal N-acyl-diacylglycerylcysteine (the cysteine being the N-terminal amino acid of the protein)[561,562].

9.1.2 Bacterial outer membrane proteins

Interestingly none of the OM components are produced at the OM: Whereas OMPs, whether they are integral or lipid-anchored, are produced in the cytoplasm, the OM phospholipids and LPS are made at the cytoplasmic side of the IM[563-565]. All these compounds therefore require transport to their final destination.

The *E. coli* genome codes for about 100 lipoproteins. The transport of OM lipoproteins has been elucidated by biochemical studies. After N-terminal lipid modification[561] at the periplasmic side of the IM they interact with the ABC transporter LolCDE. IM lipoproteins are not bound by this complex, because they possess a so-called 'Lol-avoidance signal'. Upon ATP hydrolysis the lipoprotein is released to the periplasmic chaperone LolA, which guides the lipoprotein through the periplasm to the OM, where it is handed over to the OM lipoprotein LolB, which assembles the lipoprotein finally in the OM[562].

The glycerophospholipids exchange freely between the outer leaflet of the IM and the inner leaflet of the OM; the mechanism for this reversible transport is not understood

yet. Lipopolysaccharides, on the other hand, are exported from the IM to the OM in an unidirectional manner[566]. The LptA/LptB proteins are responsible for the transport of LPS across the periplasmic space[567]. Whether this transport occurs by diffusion of the LPS loaded LptA/LptB complex through the periplasm or whether the transport is carried out during the formation of transient contact sites, is still under dispute[568,569].

After their synthesis in the cytoplasm a signal sequence targets the OMPs to cross the IM. The protein machinery involved in the translocation of bacterial membrane proteins is called the "Sec machinery", as all those components were originally identified in screens for deficiencies in protein secretion[570]. The Sec translocation machinery can be divided into two functionally different subsets: The chaperone SecB is a soluble cytoplasmic protein and keeps the proteins in a translocation competent (i.e. unfolded) state and at the same time prevents aggregation by shielding exposed hydrophobic sequences. The SecY, SecE, and SecG are all membrane proteins and interact to form a complex (SecYEG). This complex interacts with the peripheral membrane protein SecA, and it is this quaternary complex that drives the translocation of secretory proteins across the IM. The energy for this translocation is provided by the proton-motive force and ATP hydrolysis[571-573]. The signal sequence and the bound SecB chaperone target the protein to the membrane embedded SecAYEG complex. After translocation the signal sequence is cleaved in the periplasm[574]. It should be noted at this point that also bacterial IMPs are processed by the Sec machinery. The treatment of these proteins differs from the secreted and OM proteins in a way that, if a hydrophobic TM segment of an IMP is detected by the Sec channel, it will open laterally and release the hydrophobic TM segment into the surrounding phospholipid bilayer[574] in an analogous fashion as it is proposed for the synthesis of eukaryotic transmembrane proteins[575,576]. Hydrophobic TM segments as they are characteristic for α-helical membrane proteins can therefore be considered as a targeting signal to the IM, which explains their absence from the OM.

While periplasmic and secreted proteins can adopt their native conformation after translocation through the IM, OMPs are immediately bound by chaperones. The most important of these are Skp and SurA[568]. Skp binds selectively to unfolded OMPs[577,578] already while they are in progress of translocation[579]. The postulated mode of action of Skp was recently confirmed in an NMR study: Trimeric Skp forms soluble 1:1 complexes with outer membrane protein A (OmpA) from *E. coli* by burying the hydrophobic TM domain within a cavity, while keeping the non-aggregation prone

periplasmic domain freely in solution in its folded form[580]. SurA, in contrast to Skp, functions as a folding chaperone for OMPs[581,582]. Data indicate that SurA acts later in the OMP insertion pathway than Skp. A working model therefore postulates that Skp initially binds nascent OMPs to prevent their aggregation and then hands them over to SurA which assists in their delivery to the OM and concomitant folding and membrane insertion[583]. After this chaperone assisted passage through the periplasm, OMPs are inserted into the OM via a machinery called β-barrel assembly machinery or Bam complex[584,585].

9.2 β -barrel proteins

9.2.1 Occurrence of β -barrel membrane proteins

The exclusive appearance of β-barrel proteins in the OM is most likely due to the fact that β-barrels are very efficient pores and would thereby disrupt the proton gradient across the IM. Several toxins act indeed by forming β-barrels within energy-transducing membranes[586].

Whereas roughly 20 amino acids are necessary to transverse the membrane with an α-helix[5], only 8 residues are required to fully span the membrane with β-strand. For both α-helix and β-strands the effective number of residues is usually higher due to a tilt of the respective structural element relative to the membrane normal. Because the side chain of only every second residue faces the hydrophobic lipid environment, β-barrel TM segments usually don't display an especially pronounced hydrophobicity.

The structures of about 30 β-barrel membrane proteins have been solved to date. The observed numbers of TM strands ranges from 8 to 22. The quaternary structure is predominantly monomeric. One class of β-barrel MPs, the porins, is trimeric. In one case an oligomer of 8 identical subunits forms one single barrel structure[587]. All β-barrel structures solved so far display an even number of β-strands with the N- and C-termini located on the periplasmic side. To both stated characteristics there is one exception: Several autotransporter proteins expose their N-terminus on the extracellular face after threading it through β-barrel pore[588], thereby seemingly "violating" the rule that N- and C-terminus have to lay on the periplasmic side. The rule that all barrels are composed of an even number of strands was broken in 2008,

when the structure of the mitochondrial OMP VDAC-1 (voltage dependent anion channel) was solved by NMR spectroscopy[215]. The presented structure showed a 19-stranded β-barrel with the first and last strand being aligned in a parallel fashion. N- and C-termini were both lying on the periplasmic face, because the C-terminus was threaded through the pore formed by the 19-stranded barrel.

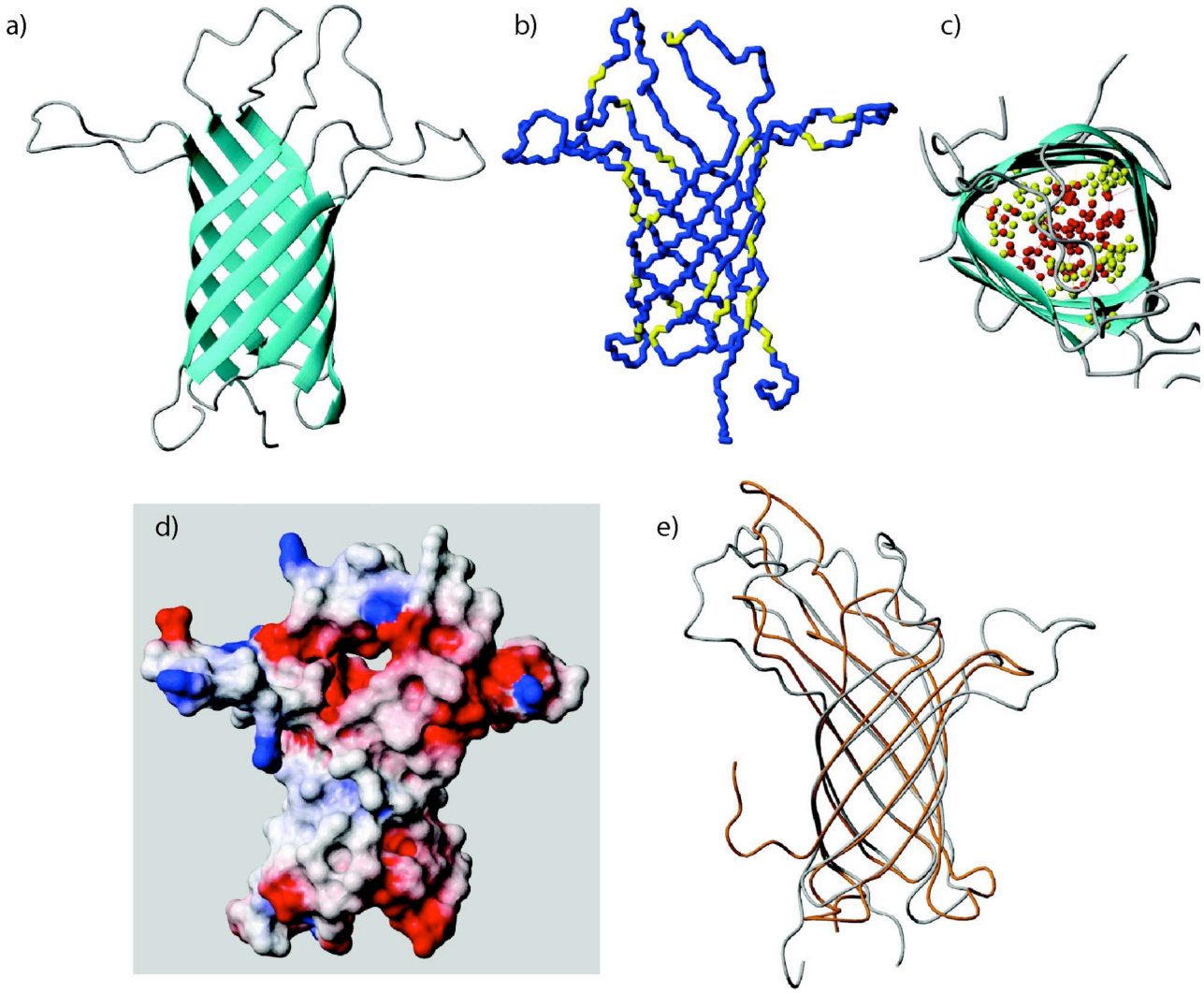


Figure 11: *a) Ribbon representation of a selected conformer of a solution NMR structure of OmpA (1G90). The β-strands of the barrel are colored cyan and delimit the surface of OmpA being inserted in the membrane. b) The characteristic aromatic belt observed in many β-barrel membrane proteins is indicated by coloring the backbone bonds of all aromatic residues of OmpA yellow. c) View from the extracellular space into the barrel of OmpA. Charged and polar residues are represented in red and yellow, respectively. d) Electrostatic surface potential of OmpA. Negatively and positively charged residues are colored and blue, respectively. The outer surface of the central, membrane embedded region is largely hydrophobic. e) A superposition of OmpA (gray; 2GE4) with a loop-shortened OmpA which had been used as a scaffold for the EF-hand loop of calmodulin (orange; 2JMM).*

9.2.2 Electrostatic characterization of β-barrel membrane proteins

Electrostatically β-barrel MPs are characterized by a hydrophobic exterior of the barrel facing the lipids of the membrane and a usually hydrophilic interior of the barrel (i.e. they show an inverted electrostatic profile compared to soluble proteins). The loops interconnecting the β-strands of the barrel are hydrophilic and short on the periplasmic and long on the extracellular side[589]. As is the case for α-helices, the antiparallel β-sheet structure fulfills the requirement of saturating all hydrogen bonding capabilities of the backbone's carbonyl oxygens and amide hydrogens[5].

The TM β-strands of β-barrel membrane proteins are rich in glycines, tryptophans and tyrosines. The two aromatic members of this trio are mainly found in the "interface" regions of the membrane.

The lumen of the smaller barrels (i.e. the eight-stranded outer membrane protein A, OmpA) is tightly packed with an interaction network of polar residues' sidechains. Pockets of ordered and unordered water molecules can also be found within these small β-barrels.

The lumen of larger β-barrels, as for instance the 12-stranded NalP from *Neisseria meningitidis*, is large enough to also accommodate secondary structure elements formed in the extracellular loops. The 16- and 18 stranded porins have pores, which are completely filled with water, and sometimes entire extracellular loops fold back into the pore, thereby modulating substrate specificity. Even larger β-barrels as the 22-stranded coupled transporter proteins contain entire "plug" or "cork" domains that fill most of the lumen.

9.3 Functions of β-barrel membrane proteins

Functionally the OM proteins of Gram-negative bacteria can be classified into six families[560]: (1) general porins, (2) passive transporters, (3) active transporters, (4) enzymes, (5) defensive proteins, and (6) structural proteins. A summary of this classification with examples and references is given in the table below.

Category	Member	Function	Ref
1. general porins	OmpC	general diffusion pore	590
	OmpF	general diffusion pore	591

	PhoE	general diffusion pore, some specificity	[591]
2. passive transporters	LamB	maltose transport	[592]
	ScrY	sugar transport	[593]
	FadL	fatty acid transport	[594]
3. active transporters	FepA	siderophore transport	[595]
	FecA	siderophore transport	[596]
	FhuA	siderophore transport	[597]
	BtuB	vitamin B12 transport	[598]
4. enzymes	OMPLA	phospholipase	[599]
	OmpT	protease	[600]
	PagP	palmitoyl transferase	[210]
5. defensive proteins	OmpX	defense	[601,602]
6. structural proteins	OmpA	structural	[214,603]

9.3.1 General porins

OmpF, OmpC and PhoE are the three major porins produced by *E. coli*. They are all trimeric, belong to the family of general porins and their crystal structures have been solved. In spite of their non-specificity they can be classified according to the charge and size of solutes which they are able to transport. While OmpF and OmpC show a slight preference for cations, PhoE prefers inorganic phosphate and anions[560,604]. While the expression of OmpF and OmpC is regulated by osmotic pressure, PhoE is expressed in response to phosphate starvation. The size limit for solutes passing through the porins is considered to be around 600 Da[605]. The general porins can play an important role in the development of antibiotic resistances, as is exemplified by the low susceptibility to β-lactams by gram-negative pathogens with a reduced number of porins in their OM[604,606].

9.3.2 Passive transporters

In the case of influx through the general porins it is assumed, that the solute does not interact with the porin, resulting in unspecific uptake and high flux rates. In case the solute interacts with its transporter, the diffusion rate will be slowed down, but specificity will be gained. The equilibrium constant of the interaction between the transporter's interior and the solute will have an influence on the flux[607]. This is the case in passive transporters such as the maltoporin LamB[608]. X-ray diffraction analysis of maltose soaked LamB crystals indicated a transport route through the

LamB channel, consisting of a row of aromatic amino acids ("greasy slide") that is flanked by polar residues ("ionic track") along which the maltose substrate is transported[609,610].

9.3.3 Active transporters

The active transporters, which are involved in the uptake of large substrates (e.g. iron siderophores or vitamin B_{12}) depend on the electrochemical potential of the cytoplasmatic membrane and an energy-transducing protein complex called TonB-ExbBD. FhuA[597] and FepA[595] are two such active transporters mediating the uptake of iron-siderophore complexes. They both form a 22-stranded β-barrel with the N-terminus forming a plug to the pore.

Enzymes

The first OM β-barrel enzyme to be characterized was the phospholipase OMPLA[611]. OMPLA is activated upon Ca^{2+}-induced dimerization[599]. Dimerization for instance is triggered by the migration of glycerophospholipids to the outer leaflet of the OM which frees Ca^{2+} ions previously involved in the bridging of the phosphates of LPS. By its phospholipase activity OMPLA is able to counteract the glycerophospholipid inflow and restore the required lipid asymmetry in the OM[612]. The active site in dimeric OMPLA is formed by a Asn-His-Ser catalytic triad located on the exterior of the β-barrel. In the monomeric state these residues are associated with a glucosamine moiety of lipid A[613].

The OM β-barrel protease OmpT was identified as a nuisance in protein purification in *E. coli*[614]. It is specific for the cleavage between two basic amino acids[615]. This specificity renders OmpT ideal in for the recognition of cationic antimicrobial peptides, in which often diads of basic amino acids can be found. Based on the crystal structure of OmpT it is expected that the O-antigen polysaccharide extends far beyond the opening cleft of OmpT and thereby shields other OMPs from digestion by OmpT[616]. Only proteins/peptides which penetrate into the O-antigen polysaccharide layer are susceptible for encounter with the OmpT active site[617].

The third OM β-barrel enzyme to be discovered is the palmitoyltransferase PagP[618]. PagP transfers a palmitate chain from the *sn*-1 position of a phospholipid to the hydroxy group of a β-hydroxyacyl chain of lipid A. This transfer occurs in the OM. PagP is an 8-stranded β-barrel as determined by X-ray crystallography[211] and NMR spectroscopy[210] and is very specific for the transfer of palmitate. This specificity is

conferred by a palmitate-recognition pocket, known as the "hydrocarbon ruler", which is located in the barrel's interior at the level of the LPS exposed half of the barrel[211].

PagP is a scavenger system to prevent accumulation of glycerophospholipids in the LPS leaflet of the OM similar to OMPLA. In contrast to OMPLA, however, PagP also functions in the absence of divalent cations and could become the only functioning rescue system under conditions of depletion of these ions[618,619]. Additionally to maintaining OM asymmetry through the degradation of glycerophospholipids the transfer of palmitoyl to lipid A also might be a strategy to avoid a host's innate immune system: after palmitoylation lipid A is less immunogenic and a bacterial invader therefore less likely to be attacked by e.g. antimicrobial peptides of the host's immune system[620].

9.3.4 Defensive proteins

The 8-stranded β-barrel OmpX is a representative of the defensive OMPs. It is important for survival of a pathogen by neutralizing a host's defense mechanism: e.g. the mediation of survival in macrophages and inhibition of the complement system[621]. The structure of OmpX has been solved both by X-ray crystallography[601] and NMR spectroscopy[602].

Structural proteins

The main structural OMP is OmpA which occurs in about 100000 copies per cell. It is composed of a 170 residue N-terminal TM domain and a 155 residue C-terminal periplasmic domain[622]. The physiological function as a structural protein consists of providing a link between the OM and the underlying peptidoglycan layer[623]. The crystal structure of the OmpA TM domain was solved in 1998[603], which was complemented by the NMR structure published in 2001[214].

9.4 Thermodynamic stability of β-barrel membrane proteins and introduction to OmpA

The outer membrane protein A (OmpA) from *E. coli* consists of a 171 residue TM domain located in the OM and a 154 residue periplasmic domain. The TM domain is an 8-stranded β-barrel and the periplasmic domain is globular[624,625]. The TM domain of OmpA could be expressed and purified independently of the periplasmic domain and its structure has been solved by X-ray crystallography[603,626] and NMR[214].

The surface exposed loops of OmpA have many functions. They are involved in the recognition of nutrients such as iron-siderophore complexes and sugars[627,628], toxic agents such as bacteriophages or colicins[629-631] and probably also in the recognition of eukaryotic targets for bacterial pathogens[632-634].

Extensive mutagenesis data showed the robustness of OmpA and its tolerance towards amino acid substitutions, especially in the longer extracellular loops, the short periplasmic turns and the face of the barrel facing the lipid bilayer[635]. The TM domain of OmpA could be circularly permutated without impairing the assembly and function in the OM[622]. The extracellular loops could be shortened to a minimal length, also without impairing the assembly and function in the OM[635]. OmpA which was denatured by chaotropic agents could successfully refolded into detergent micelles[636] and lipid bilayers[637] by rapid dilution of the denaturant. The simplicity of OmpA refolding led to the development of expression systems, for which OmpA accumulates as insoluble inclusion bodies[638]. The inclusion body expression system allows to purify denatured OmpA in large quantities (>100 mg per liter of bacterial culture).

OmpA has long been a popular model system for the study of refolding of membrane proteins. Apart from its relative abundance in the *E. coli* OM the possibility for easily monitoring its folding state by gel electrophoresis[639] have accounted for its popularity. This assay has been used to study the unfolding of urea denatured OmpA in small unilamellar vesicles. A gradual shift from the folded to the unfolded state with a midpoint at approximately 3 M urea could be observed by SDS-PAGE. Fluorescence spectroscopy experiments at pH 7 yielded the free energy of unfolding, $\Delta G^0_{u,H2O}$ to be 4.5 kcal/mol[589]. Larger β-barrels whose lumina are blocked by "plug" domains, such as FhuA, show independent unfolding of the "plug" and the β-barrel. Whereas the "plug" domain unfolds a 65 °C the β-barrel is stable up to 74 °C[640].

9.5 Mechanism of *in vitro* β-barrel membrane protein folding

The kinetics of folding of OmpA in dioleoyl-*sn*-phosphatidylcholine (DOPC) lipid bilayers proceeds in three distinct phases. The fastest phase is the association of the unfolded protein with the lipid bilayer. It has a time constant of 6 min and is independent of temperature. The second phase is the partial, but not yet complete insertion of the β-strands into the bilayer. This phase had a time constant of 15 min to 3 h and was strongly temperature dependent. The third phase marks the termination of

the folding process and ends with the complete insertion of the β-strands and formation of the tertiary structure. This last phase could only be observed at temperatures above 30 °C and occurred with a time constant of about 2 h. At 37 °C complete folding could be observed after 6 h[641]. Placing fluorescence quenchers at different depths in the membrane bilayer by covalent attachment to methylene groups in the fatty acid chains allows to measure the immersion depth of tryptophan residues in a time-resolved manner. This technique was called time-resolved distance determination by fluorescence quenching (TDFQ)[642]. Application of this technique to the folding of OmpA showed that all of the four extracellular loops crossed the membrane with the same time course[643]. Therefore OmpA inserts and folds in the membrane by a mechanism in which all extracellular loops translocate through the bilayer in a concerted manner. This movement of the loops, which are followed by the β-strands leads to a continuous insertion of the β-strands and a concomitant formation of tertiary structure. This is in contrast to the two-stage folding model suggested for α-helical membrane proteins[644] in which individual TM helices insert into the membrane independently and then in a second stage associate to form the tertiary structure.

The whole OmpA membrane insertion and folding process can be described as the hydrophobic collapse of the unfolded state U into an intermediate called I_W in solution, followed by the binding of I_W to the membrane surface, where intermediate I_{M1}, which is characterized by disordered Trps, is formed. No secondary structure is observed in I_{M1}. Subsequently intermediate I_{M2} is formed, with all Trps located at a distance of about 10 Å from the membrane center. A fraction of the secondary structure elements is formed in this intermediate, which is sometimes also referred to as a "molten disk". Next the β-strands start to move towards the membrane center, and in intermediate I_{M3} the Trps located near the extracellular loops are located in the membrane center, whereas a Trp located near the periplasmic side remains at a distance of 10 Å from the center. This structure is more globular and is therefore also referred to as "molten globule". Secondary structural elements are formed in this state, but the correct tertiary fold has not been reached yet. In the last step I_{M3} proceeds to form the fully folded native state N, in which the Trps located near the extracellular loops are again located at a distance of about 10 Å from the membrane center, but this time on the side of the membrane opposite to where they started their

journey.

9.6 *In vivo* folding of β -barrel membrane proteins

In vivo the folding of OMPs is much faster than *in vitro*[645] and it is therefore likely that there are proteins assisting the efficient membrane incorporation of OMPs. The two main periplasmic chaperones are Skp and SurA (*vide supra*). Skp seems to be an ideal shuttling protein, because it is soluble, but can bind peripherally to phospholipid bilayers[578]. It is able to bind many different nascent OMPs[577]. However, it has been shown, that Skp does not accelerate folding of its substrates. It therefore seems to be a shuttle rather than an active folding catalyst.

SurA is a periplasmic peptidyl-prolyl isomerase that assists the folding of several OMPs[581].

10. References

1. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63**, 1256-1272 (2003).
2. Fadeel, B. & Xue, D. The ins and outs of phospholipid asymmetry in the plasma membrane: roles in health and disease. *Crit Rev Biochem Mol Biol* **44**, 264-277 (2009).
3. Lewis, B. A. & Engelman, D. M. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *J Mol Biol* **166**, 211-217 (1983).
4. Wiener, M. C. & White, S. H. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. II. Distribution and packing of terminal methyl groups. *Biophys J* **61**, 428-433 (1992).
5. White, S. H., Ladokhin, A. S., Jayasinghe, S. & Hristova, K. How membranes shape protein structure. *J Biol Chem* **276**, 32395-32398 (2001).
6. Killian, J. A. & von Heijne, G. How proteins adapt to a membrane-water interface. *Trends Biochem Sci* **25**, 429-434 (2000).
7. White, S. H. & Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**, 319-365 (1999).
8. Magee, T. & Seabra, M. C. Fatty acylation and prenylation of proteins: what's hot in fat. *Curr Opin Cell Biol* **17**, 190-196 (2005).
9. Roskoski, R. J. Protein prenylation: a pivotal posttranslational process. *Biochem Biophys Res Commun* **303**, 1-7 (2003).
10. Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G. & Eisenhaber, F. Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins. *Bioessays* **25**, 367-385 (2003).
11. Takeda, J. & Kinoshita, T. GPI-anchor biosynthesis. *Trends Biochem Sci* **20**, 367-371 (1995).
12. Fernandez, C., Hilty, C., Wider, G. & Wüthrich, K. Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy. *Proc Natl Acad Sci U S A* **99**, 13533-13537 (2002).
13. Lee, D. et al. Bilayer in small bicelles revealed by lipid-protein interactions using NMR spectroscopy. *J Am Chem Soc* **130**, 13822-13823 (2008).

14. Neumoin, A., Arshava, B., Becker, J., Zerbe, O. & Naider, F. NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae. *Biophys J* **93**, 467-482 (2007).
15. Dale, H. H. On some physiological actions of ergot. *J Physiol* **34**, 163-206 (1906).
16. Langley, J. N. Observations on the physiological action of extracts of the supra-renal bodies. *J Physiol* **27**, 237-256 (1901).
17. Rodbell, M., Birnbaumer, L., Pohl, S. L. & Krans, H. M. The glucagon-sensitive adenyl cyclase system in plasma membranes of rat liver. V. An obligatory role of guanylnucleotides in glucagon action. *J Biol Chem* **246**, 1877-1882 (1971).
18. Sutherland, E. W. Cyclic AMP. In An Introduction. 5-13 (1971).
19. Walsh, D. A., Perkins, J. P. & Krebs, E. G. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *J Biol Chem* **243**, 3763-3765 (1968).
20. Benovic, J. L., Shorr, R. G., Caron, M. G. & Lefkowitz, R. J. The mammalian beta 2-adrenergic receptor: purification and characterization. *Biochemistry* **23**, 4510-4518 (1984).
21. Dixon, R. A. et al. Cloning of the gene and cDNA for mammalian beta-adrenergic receptor and homology with rhodopsin. *Nature* **321**, 75-79 (1986).
22. Hargrave, P. A. et al. The structure of bovine rhodopsin. *Biophys Struct Mech* **9**, 235-244 (1983).
23. Ovchinnikov, Y. Rhodopsin and bacteriorhodopsin: structure-function relationships. *FEBS Lett* **148**, 179-191 (1982).
24. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-187 (1991).
25. Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **3**, 928-944 (2002).
26. Lee, D. K., George, S. R. & O'Dowd, B. F. Continued discovery of ligands for G protein-coupled receptors. *Life Sci* **74**, 293-297 (2003).
27. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
28. Stryer, L. Cyclic GMP cascade of vision. *Annu Rev Neurosci* **9**, 87-119 (1986).
29. Sunahara, R. K., Dessauer, C. W., Whisnant, R. E., Kleuss, C. & Gilman, A. G. Interaction of Gsalpha with the cytosolic domains of mammalian adenylyl cyclase. *J Biol Chem* **272**, 22265-22271 (1997).
30. Downes, G. B. & Gautam, N. The G protein subunit gene families. *Genomics* **62**, 544-552 (1999).
31. Cabrera-Vera, T. M. et al. Insights into G protein structure, function, and regulation. *Endocr Rev* **24**, 765-781 (2003).
32. Riobo, N. A. & Manning, D. R. Receptors coupled to heterotrimeric G proteins of the G12 family. *Trends Pharmacol Sci* **26**, 146-154 (2005).
33. Garcia, P. D., Onrust, R., Bell, S. M., Sakmar, T. P. & Bourne, H. R. Transducin-alpha C-terminal mutations prevent activation by rhodopsin: a new assay using recombinant proteins expressed in cultured cells. *EMBO J* **14**, 4460-4469 (1995).
34. Hamm, H. E. et al. Site of G protein binding to rhodopsin mapped with synthetic peptides from the alpha subunit. *Science* **241**, 832-835 (1988).
35. Dratz, E. A. et al. NMR structure of a receptor-bound G-protein peptide. *Nature* **363**, 276-281 (1993).
36. Onrust, R. et al. Receptor and betagamma binding sites in the alpha subunit of the retinal G protein transducin. *Science* **275**, 381-384 (1997).
37. Romano, M., Maddox, J. F. & Serhan, C. N. Activation of human monocytes and the acute monocytic leukemia cell line (THP-1) by lipoxins involves unique signaling pathways for lipoxin A4 versus lipoxin B4: evidence for differential Ca2+ mobilization. *J Immunol* **157**, 2149-2154 (1996).
38. Klein, C. et al. Identification of surrogate agonists for the human FPRL-1 receptor by autocrine selection in yeast. *Nat Biotechnol* **16**, 1334-1337 (1998).
39. Schertler, G. F. X. Overproduction of membrane proteins. *Curr. Opin. Struct. Biol* **2**, 534 (1992).
40. Zack, D. J. et al. Unusual topography of bovine rhodopsin promoter-lacZ fusion gene expression in transgenic mouse retinas. *Neuron* **6**, 187-199 (1991).
41. Filipek, S., Teller, D. C., Palczewski, K. & Stenkamp, R. The crystallographic model of rhodopsin and its use in studies of other G protein-coupled receptors. *Annu Rev Biophys Biomol Struct* **32**, 375-397 (2003).
42. Koulen, P. & Brandstatter, J. H. Pre- and Postsynaptic Sites of Action of mGluR8a in the

mammalian retina. *Invest Ophthalmol Vis Sci* **43**, 1933-1940 (2002).
43. Pootanakit, K., Prior, K. J., Hunter, D. D. & Brunken, W. J. 5-HT2a receptors in the rabbit retina: potential presynaptic modulators. *Vis Neurosci* **16**, 221-230 (1999).
44. Straiker, A. et al. Cannabinoid CB1 receptors and ligands in vertebrate retina: localization and function of an endogenous signaling system. *Proc Natl Acad Sci U S A* **96**, 14565-14570 (1999).
45. Tai, A. W., Chuang, J. Z., Bode, C., Wolfrum, U. & Sung, C. H. Rhodopsin's carboxy-terminal cytoplasmic tail acts as a membrane receptor for cytoplasmic dynein by binding to the dynein light chain Tctex-1. *Cell* **97**, 877-887 (1999).
46. Grisshammer, R. & Tate, C. G. Overexpression of integral membrane proteins for structural studies. *Q Rev Biophys* **28**, 315-422 (1995).
47. van der Vliet, P. C. Special Issue on Overexpression of Integral Membrane Proteins. *Biochim Biophys Acta* **1610**, 1-153
48. Grisshammer, R., White, J. F., Trinh, L. B. & Shiloach, J. Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J Struct Funct Genomics* **6**, 159-163 (2005).
49. Krepkiy, D., Wong, K., Gawrisch, K. & Yeliseev, A. Bacterial expression of functional, biotinylated peripheral cannabinoid receptor CB2. *Protein Expr Purif* **49**, 60-70 (2006).
50. LaVallie, E. R. & McCoy, J. M. Gene fusion expression systems in Escherichia coli. *Curr Opin Biotechnol* **6**, 501-506 (1995).
51. Weiss, H. M. & Grisshammer, R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur J Biochem* **269**, 82-92 (2002).
52. White, J. F., Trinh, L. B., Shiloach, J. & Grisshammer, R. Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett* **564**, 289-293 (2004).
53. Tucker, J. & Grisshammer, R. Purification of a rat neurotensin receptor expressed in Escherichia coli. *Biochem J* **317**, 891-899 (1996).
54. Warne, T., Chirnside, J. & Schertler, G. F. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim Biophys Acta* **1610**, 133-140 (2003).
55. Booth, P. J. The trials and tribulations of membrane protein folding in vitro. *Biochim Biophys Acta* **1610**, 51-56 (2003).
56. Baneres, J. L. et al. Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1. *J Mol Biol* **329**, 801-814 (2003).
57. Loll, P. J. Membrane protein structural biology: the high throughput challenge. *J Struct Biol* **142**, 144-153 (2003).
58. Gimpl, G., Burger, K. & Fahrenholz, F. Cholesterol as modulator of receptor function. *Biochemistry* **36**, 10959-10974 (1997).
59. Lilie, H., Schwarz, E. & Rudolph, R. Advances in refolding of proteins produced in E. coli. *Curr Opin Biotechnol* **9**, 497-501 (1998).
60. Laage, R. & Langosch, D. Strategies for prokaryotic expression of eukaryotic membrane proteins. *Traffic* **2**, 99-104 (2001).
61. Wacker, M. et al. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **298**, 1790-1793 (2002).
62. Kowarik, M. et al. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148-1150 (2006).
63. Slynko, V. et al. NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc* **131**, 1274-1281 (2009).
64. David, N. E. et al. Expression and purification of the Saccharomyces cerevisiae alpha-factor receptor (Ste2p), a 7-transmembrane-segment G protein-coupled receptor. *J Biol Chem* **272**, 15553-15561 (1997).
65. Reiländer, H. & Weiss, H. M. Production of G-protein-coupled receptors in yeast. *Curr Opin Biotechnol* **9**, 510-517 (1998).
66. Werten, M. W., van den Bosch, T. J., Wind, R. D., Mooibroek, H. & de Wolf, F. A. High-yield secretion of recombinant gelatins by Pichia pastoris. *Yeast* **15**, 1087-1096 (1999).
67. King, K., Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. Control of yeast mating signal transduction by a mammalian beta 2-adrenergic receptor and Gs alpha subunit. *Science* **250**, 121-123 (1990).
68. Crespi, H. L., Rosenberg, R. M. & Katz, J. J. Proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains. *Science* **161**, 795-796 (1968).
69. Katz, J. J. & Crespi, H. L. Deuterated organisms: cultivation and uses. *Science* **151**, 1187-1194

(1966).
70. Gardner, K. H. & Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**, 357-406 (1998).
71. Morgan, W. D., Kragt, A. & Feeney, J. Expression of deuterium-isotope-labelled protein in the yeast pichia pastoris for NMR studies. *J Biomol NMR* **17**, 337-347 (2000).
72. Hamilton, S. R. & Gerngross, T. U. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* **18**, 387-392 (2007).
73. Smith, G. E., Summers, M. D. & Fraser, M. J. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol* **3**, 2156-2165 (1983).
74. Lynn, D. E. Available Lepidopteran insect cell lines. *Baculovirus and Insect Cell Expression Protocols* (2007).
75. Jarvis, D. L. & Finn, E. E. Biochemical analysis of the N-glycosylation pathway in baculovirus-infected lepidopteran insect cells. *Virology* **212**, 500-511 (1995).
76. Grünewald, S., Haase, W., Reiländer, H. & Michel, H. Glycosylation, palmitoylation, and localization of the human D2S receptor in baculovirus-infected insect cells. *Biochemistry* **35**, 15149-15161 (1996).
77. Massotte, D., Pereira, C. A., Pouliquen, Y. & Pattus, F. Parameters influencing human mu opioid receptor over-expression in baculovirus-infected insect cells. *J Biotechnol* **69**, 39-45 (1999).
78. Leifert, W. R., Aloia, A. L., Bucco, O., Glatz, R. V. & McMurchie, E. J. G-protein-coupled receptors in drug discovery: nanosizing using cell-free technologies and molecular biology approaches. *J Biomol Screen* **10**, 765-779 (2005).
79. Massotte, D. G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta* **1610**, 77-89 (2003).
80. Fraser, M. J. The baculovirus-infected insect cell as a eukaryotic gene expression system. *Curr Top Microbiol Immunol* **158**, 131-172 (1992).
81. Green, T., Stauffer, K. A. & Lummis, S. C. Expression of recombinant homo-oligomeric 5-hydroxytryptamine3 receptors provides new insights into their maturation and structure. *J Biol Chem* **270**, 6056-6061 (1995).
82. Ratnala, V. R. et al. Large-scale overproduction, functional purification and ligand affinities of the His-tagged human histamine H1 receptor. *Eur J Biochem* **271**, 2636-2646 (2004).
83. Ruan, K. H., Cervantes, V. & Wu, J. A simple, quick, and high-yield preparation of the human thromboxane A2 receptor in full size for structural studies. *Biochemistry* **47**, 6819-6826 (2008).
84. Robeva, A. S., Woodard, R., Luthin, D. R., Taylor, H. E. & Linden, J. Double tagging recombinant A1- and A2A-adenosine receptors with hexahistidine and the FLAG epitope. Development of an efficient generic protein purification procedure. *Biochem Pharmacol* **51**, 545-555 (1996).
85. Murphree, L. J., Marshall, M. A., Rieger, J. M., MacDonald, T. L. & Linden, J. Human A(2A) adenosine receptors: high-affinity agonist binding to receptor-G protein complexes containing Gbeta(4). *Mol Pharmacol* **61**, 455-462 (2002).
86. Glass, M. & Northup, J. K. Agonist selective regulation of G proteins by cannabinoid CB(1) and CB(2) receptors. *Mol Pharmacol* **56**, 1362-1369 (1999).
87. Akermoun, M. et al. Characterization of 16 human G protein-coupled receptors expressed in baculovirus-infected insect cells. *Protein Expr Purif* **44**, 65-74 (2005).
88. Eroglu, C., Cronet, P., Panneels, V., Beaufils, P. & Sinning, I. Functional reconstitution of purified metabotropic glutamate receptor expressed in the fly eye. *EMBO Rep* **3**, 491-496 (2002).
89. Zhang, L. et al. Expression of functional G protein-coupled receptors in photoreceptors of transgenic Xenopus laevis. *Biochemistry* **44**, 14509-14518 (2005).
90. Kassack, M. U. et al. Functional screening of G protein-coupled receptors by measuring intracellular calcium with a fluorescence microplate reader. *J Biomol Screen* **7**, 233-246 (2002).
91. Pupo, A. S. & Minneman, K. P. Interaction of neuronal nitric oxide synthase with alpha1-adrenergic receptor subtypes in transfected HEK-293 cells. *BMC Pharmacol* **2**, 17 (2002).
92. Xu, J. et al. Heterodimerization of alpha 2A- and beta 1-adrenergic receptors. *J Biol Chem* **278**, 10770-10777 (2003).
93. Reeves, P. J. et al. Expression and purification of rhodopsin and its mutants from stable mammalian cell lines: application to NMR studies. *Biochem Soc Trans* **27**, 950-955 (1999).
94. Yin, D. et al. Successful expression of a functional yeast G-protein-coupled receptor (Ste2) in mammalian cells. *Biochem Biophys Res Commun* **329**, 281-287 (2005).
95. Oprian, D. D., Molday, R. S., Kaufman, R. J. & Khorana, H. G. Expression of a synthetic

bovine rhodopsin gene in monkey kidney cells. *Proc Natl Acad Sci U S A* **84**, 8874-8878 (1987).
96. Sun, T., Naini, A. A. & Miller, C. High-level expression and functional reconstitution of Shaker K+ channels. *Biochemistry* **33**, 9992-9999 (1994).
97. Saudou, F., Boschert, U., Amlaiky, N., Plassat, J. L. & Hen, R. A family of Drosophila serotonin receptors with distinct intracellular signalling properties and expression patterns. *EMBO J* **11**, 7-17 (1992).
98. Lundstrom, K. Alphaviruses as tools in neurobiology and gene therapy. *J Recept Signal Transduct Res* **19**, 673-686 (1999).
99. Lundstrom, K. Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim Biophys Acta* **1610**, 90-96 (2003).
100. Lundstrom, K. et al. Semliki Forest virus vectors: efficient vehicles for in vitro and in vivo gene delivery. *FEBS Lett* **504**, 99-103 (2001).
101. Lohse, M. J. Stable overexpression of human beta 2-adrenergic receptors in mammalian cells. *Naunyn Schmiedebergs Arch Pharmacol* **345**, 444-451 (1992).
102. Hansen, A. P. et al. A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. *Biochemistry* **31**, 12713-12718 (1992).
103. Swartz, J. R. Universal cell-free protein synthesis. *Nat Biotechnol* **27**, 731-732 (2009).
104. Lian, L. Y. Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 171 (2001).
105. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol* **177**, 44-73 (1989).
106. Elbaz, Y., Steiner-Mordoch, S., Danieli, T. & Schuldiner, S. In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proc Natl Acad Sci U S A* **101**, 1519-1524 (2004).
107. Klammt, C. et al. High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* **271**, 568-580 (2004).
108. Ishihara, G. et al. Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein Expr Purif* **41**, 27-37 (2005).
109. Klammt, C. et al. Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J* **272**, 6024-6038 (2005).
110. Kalmbach, R. et al. Functional cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes. *J Mol Biol* **371**, 639-648 (2007).
111. Nozawa, A. et al. A cell-free translation and proteoliposome reconstitution system for functional analysis of plant solute transporters. *Plant Cell Physiol* **48**, 1815-1820 (2007).
112. Wuu, J. J. & Swartz, J. R. High yield cell-free production of integral membrane proteins without refolding or detergents. *Biochim Biophys Acta* **1778**, 1237-1250 (2008).
113. Madin, K., Sawasaki, T., Ogasawara, T. & Endo, Y. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci U S A* **97**, 559-564 (2000).
114. Zubay, G. In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7**, 267-287 (1973).
115. Pornillos, O., Chen, Y. J., Chen, A. P. & Chang, G. X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* **310**, 1950-1953 (2005).
116. Shimada, Y., Wang, Z. Y., Mochizuki, Y., Kobayashi, M. & Nozawa, T. Functional expression and characterization of a bacterial light-harvesting membrane protein in Escherichia coli and cell-free synthesis systems. *Biosci Biotechnol Biochem* **68**, 1942-1948 (2004).
117. Kigawa, T. et al. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* **442**, 15-19 (1999).
118. Liu, D. V., Zawada, J. F. & Swartz, J. R. Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnol Prog* **21**, 460-465 (2005).
119. Yamane, T., Ikeda, Y., Nagasaka, T. & Nakano, H. Enhanced cell-free protein synthesis using a S30 extract from Escherichia coli grown rapidly at 42 degrees C in an amino acid enriched medium. *Biotechnol Prog* **21**, 608-613 (2005).
120. Sawasaki, T. et al. The wheat germ cell-free expression system: methods for high-throughput materialization of genetic information. *Methods Mol Biol* **310**, 131-144 (2005).
121. Shirokov, V. A. Conitnuous-exchange protein-synthesiszing system. *In Vitro Transcription and Translation Protocols, 2nd Ed.*
122. Kamonchanok, S. et al. GPCR proteomics: mass spectrometric and functional analysis of histamine H1 receptor after baculovirus-driven and in vitro cell free expression. *J Proteome Res*

7, 621-629 (2008).
123. Klammt, C. et al. Cell-free production of G protein-coupled receptors for functional and structural studies. *J Struct Biol* **158**, 482-493 (2007).
124. Kim, D. M. & Swartz, J. R. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol Bioeng* **66**, 180-188 (1999).
125. Klammt, C. et al. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS J* **273**, 4141-4153 (2006).
126. Keppetipola, S. et al. *From gene to HSQC in under five hours: High-throughput NMR proteomics* Univ Calif Irvine, Dept Chem, Irvine, CA 92697 USA ajshaka@uci.edu, 2006).
127. Ozawa, K., Wu, P. S., Dixon, N. E. & Otting, G. N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *FEBS J* **273**, 4154-4159 (2006).
128. Albert, A. D. & Litman, B. J. Independent structural domains in the membrane protein bovine rhodopsin. *Biochemistry* **17**, 3893-3900 (1978).
129. Schertler, G. F., Villa, C. & Henderson, R. Projection structure of rhodopsin. *Nature* **362**, 770-772 (1993).
130. Okada, T. et al. X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J Struct Biol* **130**, 73-80 (2000).
131. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
132. Rasmussen, S. G. et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383-387 (2007).
133. Rosenbaum, D. M. et al. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**, 1266-1273 (2007).
134. Cherezov, V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
135. Hanson, M. A. et al. A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure* **16**, 897-905 (2008).
136. Roth, C. B., Hanson, M. A. & Stevens, R. C. Stabilization of the human beta2-adrenergic receptor TM4-TM3-TM5 helix interface by mutagenesis of Glu122(3.41), a critical residue in GPCR structure. *J Mol Biol* **376**, 1305-1319 (2008).
137. Warne, T. et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491 (2008).
138. Jaakola, V. P. et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211-1217 (2008).
139. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183-187 (2008).
140. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci U S A* **105**, 877-882 (2008).
141. Probst, W. C., Snyder, L. A., Schuster, D. I., Brosius, J. & Sealfon, S. C. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol* **11**, 1-20 (1992).
142. Applebury, M. L. & Hargrave, P. A. Molecular biology of the visual pigments. *Vision Res* **26**, 1881-1895 (1986).
143. Shi, L. et al. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J Biol Chem* **277**, 40989-40996 (2002).
144. Ballesteros, J. A. et al. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* **276**, 29171-29177 (2001).
145. Shapiro, D. A., Kristiansen, K., Weiner, D. M., Kroeze, W. K. & Roth, B. L. Evidence for a model of agonist-induced activation of 5-hydroxytryptamine 2A serotonin receptors that involves the disruption of a strong ionic interaction between helices 3 and 6. *J Biol Chem* **277**, 11441-11449 (2002).
146. Vogel, R. et al. Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. *J Mol Biol* **380**, 648-655 (2008).
147. Hanson, M. A. & Stevens, R. C. Discovery of new GPCR biology: one receptor structure at a time. *Structure* **17**, 8-14 (2009).
148. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356-363 (2009).
149. Fritze, O. et al. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and

during activation. *Proc Natl Acad Sci U S A* **100**, 2290-2295 (2003).
150. Yao, X. et al. Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. *Nat Chem Biol* **2**, 417-422 (2006).
151. Nakamichi, H., Buss, V. & Okada, T. Photoisomerization mechanism of rhodopsin and 9-cis-rhodopsin revealed by x-ray crystallography. *Biophys J* **92**, L106-8 (2007).
152. Nakamichi, H. & Okada, T. Local peptide movement in the photoreaction intermediate of rhodopsin. *Proc Natl Acad Sci U S A* **103**, 12729-12734 (2006).
153. Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C. & Schertler, G. F. Electron crystallography reveals the structure of metarhodopsin I. *EMBO J* **23**, 3609-3620 (2004).
154. Salom, D. et al. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc Natl Acad Sci U S A* **103**, 16123-16128 (2006).
155. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363-367 (2008).
156. Shimamura, T. et al. Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J Biol Chem* **283**, 17753-17756 (2008).
157. Sakmar, T. P., Franke, R. R. & Khorana, H. G. Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc Natl Acad Sci U S A* **86**, 8309-8313 (1989).
158. Arnis, S., Fahmy, K., Hofmann, K. P. & Sakmar, T. P. A conserved carboxylic acid group mediates light-dependent proton uptake and signaling by rhodopsin. *J Biol Chem* **269**, 23879-23881 (1994).
159. Mirzadegan, T., Benko, G., Filipek, S. & Palczewski, K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* **42**, 2759-2767 (2003).
160. Okada, T. et al. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. *J Mol Biol* **342**, 571-583 (2004).
161. Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **343**, 1409-1438 (2004).
162. Scheerer, P. et al. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497-502 (2008).
163. Lamb, T. D. & Pugh, E. N. J. Dark adaptation and the retinoid cycle of vision. *Prog Retin Eye Res* **23**, 307-380 (2004).
164. Vogel, R. & Siebert, F. Conformations of the active and inactive states of opsin. *J Biol Chem* **276**, 38487-38493 (2001).
165. Cohen, G. B., Oprian, D. D. & Robinson, P. R. Mechanism of activation and inactivation of opsin: role of Glu113 and Lys296. *Biochemistry* **31**, 12592-12601 (1992).
166. Farahbakhsh, Z. T., Ridge, K. D., Khorana, H. G. & Hubbell, W. L. Mapping light-dependent structural changes in the cytoplasmic loop connecting helices C and D in rhodopsin: a site-directed spin labeling study. *Biochemistry* **34**, 8812-8819 (1995).
167. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **274**, 768-770 (1996).
168. Lagerstrom, M. C. & Schioth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **7**, 339-357 (2008).
169. Sun, C. et al. Solution structure and mutational analysis of pituitary adenylate cyclase-activating polypeptide binding to the extracellular domain of PAC1-RS. *Proc Natl Acad Sci U S A* **104**, 7875-7880 (2007).
170. Pioszak, A. A. & Xu, H. E. Molecular recognition of parathyroid hormone by its G protein-coupled receptor. *Proc Natl Acad Sci U S A* **105**, 5034-5039 (2008).
171. Parthier, C. et al. Crystal structure of the incretin-bound extracellular domain of a G protein-coupled receptor. *Proc Natl Acad Sci U S A* **104**, 13942-13947 (2007).
172. Runge, S., Thogersen, H., Madsen, K., Lau, J. & Rudolph, R. Crystal structure of the ligand-bound glucagon-like peptide-1 receptor extracellular domain. *J Biol Chem* **283**, 11340-11347 (2008).
173. Grace, C. R. et al. NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proc Natl Acad Sci U S A* **101**, 12836-12841 (2004).
174. Grace, C. R. et al. Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand. *Proc Natl Acad Sci U S A* **104**, 4858-4863 (2007).
175. Grace, C. R. et al. NMR structure of the first extracellular domain of corticotropin-releasing factor receptor 1 (ECD1-CRF-R1) complexed with a high affinity agonist. *J Biol Chem* **285**, 38580-38589 (2010).

176. Pioszak, A. A., Parker, N. R., Suino-Powell, K. & Xu, H. E. Molecular recognition of corticotropin-releasing factor by its G-protein-coupled receptor CRFR1. *J Biol Chem* **283**, 32900-32912 (2008).
177. Hoare, S. R. Mechanisms of peptide and nonpeptide ligand binding to Class B G-protein-coupled receptors. *Drug Discov Today* **10**, 417-427 (2005).
178. Kolakowski, L. F. Gcrdb - a G-Protein-Coupled Receptor Database. *Receptors & Channels* **2**, 1-7 (1994).
179. Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* **18**, 1723-1729 (1999).
180. Ulrich, C. D. n., Holtmann, M. & Miller, L. J. Secretin and vasoactive intestinal peptide receptors: members of a unique family of G protein-coupled receptors. *Gastroenterology* **114**, 382-397 (1998).
181. O'Hara, P. J. et al. The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41-52 (1993).
182. Conn, P. J. & Pin, J. P. Pharmacology and functions of metabotropic glutamate receptors. *Annu Rev Pharmacol Toxicol* **37**, 205-237 (1997).
183. Baldwin, J. M., Schertler, G. F. & Unger, V. M. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol* **272**, 144-164 (1997).
184. Schwartz, T. W., Gether, U., Scahmbye, H. T. & Hjorth, S. A. Molecular Mechanism of Action of Non-peptide Ligands for Peptide Receptors. *Curr. Pharmaceutical Design* **1**, 325-342 (1995).
185. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods in Neurosciences* **25**, 366-428 (1995).
186. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-920 (2000).
187. Yusupov, M. M. et al. Crystal structure of the ribosome at 5.5 A resolution. *Science* **292**, 883-896 (2001).
188. Schuwirth, B. S. et al. Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310**, 827-834 (2005).
189. Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 A resolution. *Science* **311**, 1263-1267 (2006).
190. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322 (2008).
191. Dixon, L. R., Markiewicz, W. D., Brey, W. W. & Shetty, K. K. *Performance of the ultra wide bore 900 MHz NMR magnet at the national high magnetic field laboratory* Natl High Magnet Field Lab, Tallahassee, FL 32310 USA dixon@magnet.fsu.edu, 2005).
192. Kovacs, H., Moskau, D. & Spraul, M. *Cryogenically cooled probes - a leap in NMR technology* Bruker Biospin AG, CH-8117 Fallanden, Switzerland helena.kovacs@bruker-biospin.ch, 2005).
193. Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L. & Sweedler, J. V. High-Resolution Microcoil 1H-NMR for Mass-Limited, Nanoliter-Volume Samples. *Science* **270**, 1967-1970 (1995).
194. Wüthrich, K. NMR of Proteins and Nucleic Acids. (1986).
195. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium. *Journal of magnetic resonance* (1992).
196. Tugarinov, V. & Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868-13878 (2003).
197. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* **1**, 749-754 (2006).
198. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-12371 (1997).
199. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* **95**, 13585-13590 (1998).
200. Tugarinov, V., Choy, W. Y., Orekhov, V. Y. & Kay, L. E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci U S A* **102**, 622-627 (2005).
201. Gao, F. P. & Cross, T. A. Recent developments in membrane-protein structural genomics.

Genome Biol **6**, 244 (2005).
202. Sanders, C. R. & Sonnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-40 (2006).
203. Columbus, L. et al. Expression, purification, and characterization of Thermotoga maritima membrane proteins for structure determination. *Protein Sci* **15**, 961-975 (2006).
204. Krueger-Koplin, R. D. et al. An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* **28**, 43-57 (2004).
205. Page, R. C. et al. Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J Struct Funct Genomics* **7**, 51-64 (2006).
206. Vinogradova, O., Sonnichsen, F. & Sanders, C. R. n. On choosing a detergent for solution NMR studies of membrane proteins. *J Biomol NMR* **11**, 381-386 (1998).
207. Zhou, Y. & Bowie, J. U. Building a thermostable membrane protein. *J Biol Chem* **275**, 6975-6979 (2000).
208. Oxenoid, K., Kim, H. J., Jacob, J., Sonnichsen, F. D. & Sanders, C. R. NMR assignments for a helical 40 kDa membrane protein. *J Am Chem Soc* **126**, 5048-5049 (2004).
209. Van Horn, W. D. et al. Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* **324**, 1726-1729 (2009).
210. Hwang, P. M. et al. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci U S A* **99**, 13560-13565 (2002).
211. Ahn, V. E. et al. A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J* **23**, 2931-2941 (2004).
212. Hwang, P. M., Bishop, R. E. & Kay, L. E. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc Natl Acad Sci U S A* **101**, 9618-9623 (2004).
213. Hawkins, C. A., de Alba, E. & Tjandra, N. Solution structure of human saposin C in a detergent environment. *J Mol Biol* **346**, 1381-1392 (2005).
214. Arora, A., Abildgaard, F., Bushweller, J. H. & Tamm, L. K. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* **8**, 334-338 (2001).
215. Hiller, S. et al. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206-1210 (2008).
216. Hauser, H. Short-chain phospholipids as detergents. *Biochim Biophys Acta* **1508**, 164-181 (2000).
217. Lichtenberg, D., Robson, R. J. & Dennis, E. A. Solubilization of phospholipids by detergents. Structural and kinetic aspects. *Biochim Biophys Acta* **737**, 285-304 (1983).
218. Lin, T. L., Chen, S. H., Gabriel, N. E. & Roberts, M. F. The use of small-angle neutron scattering to determine the structure and interaction of dihexanoylphosphatidylcholine micelles. *J Am Chem Soc* **108**, 3499-3507 (1986).
219. Rosenbusch, J. P. The critical role of detergents in the crystallization of membrane proteins. *J Struct Biol* **104**, 134-138 (1990).
220. Prosser, R. S., Evanics, F., Kitevski, J. L. & Al-Abdul-Wahid, M. S. Current applications of bicelles in NMR studies of membrane-associated amphiphiles and proteins. *Biochemistry* **45**, 8453-8465 (2006).
221. De Angelis, A. A., Howell, S. C., Nevzorov, A. A. & Opella, S. J. Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy. *J Am Chem Soc* **128**, 12256-12267 (2006).
222. Park, S. H. et al. High-resolution NMR spectroscopy of a GPCR in aligned bicelles. *J Am Chem Soc* **128**, 7402-7403 (2006).
223. Park, S. H., De Angelis, A. A., Nevzorov, A. A., Wu, C. H. & Opella, S. J. Three-dimensional structure of the transmembrane domain of Vpu from HIV-1 in aligned phospholipid bicelles. *Biophys J* **91**, 3032-3042 (2006).
224. Dürr, U. H., Yamamoto, K., Im, S. C., Waskell, L. & Ramamoorthy, A. Solid-state NMR reveals structural and dynamical properties of a membrane-anchored electron-carrier protein, cytochrome b5. *J Am Chem Soc* **129**, 6670-6671 (2007).
225. Triba, M. N., Zoonens, M., Popot, J. L., Devaux, P. F. & Warschawski, D. E. Reconstitution and alignment by a magnetic field of a beta-barrel membrane protein in bicelles. *Eur Biophys J* **35**, 268-275 (2006).
226. Lyukmanova, E. N. et al. Lipid-protein nanoscale bilayers: a versatile medium for NMR investigations of membrane proteins and membrane-active peptides. *J Am Chem Soc* **130**, 2140-2141 (2008).
227. Raschle, T. et al. Structural and functional characterization of the integral membrane protein

VDAC-1 in lipid bilayer nanodiscs. *J Am Chem Soc* **131**, 17777-17779 (2009).
228. Zoonens, M., Catoire, L. J., Giusti, F. & Popot, J. L. NMR study of a membrane protein in detergent-free aqueous solution. *Proc Natl Acad Sci U S A* **102**, 8893-8898 (2005).
229. Wand, A. J., Ehrhardt, M. R. & Flynn, P. F. High-resolution NMR of encapsulated proteins dissolved in low-viscosity fluids. *Proc Natl Acad Sci U S A* **95**, 15299-15302 (1998).
230. Babu, C. R., Flynn, P. F. & Wand, A. J. Preparation, characterization, and NMR spectroscopy of encapsulated proteins dissolved in low viscosity fluids. *J Biomol NMR* **25**, 313-323 (2003).
231. Shi, Z., Peterson, R. W. & Wand, A. J. New reverse micelle surfactant systems optimized for high-resolution NMR spectroscopy of encapsulated proteins. *Langmuir* **21**, 10632-10637 (2005).
232. Peterson, R. W., Lefebvre, B. G. & Wand, A. J. High-resolution NMR studies of encapsulated proteins in liquid ethane. *J Am Chem Soc* **127**, 10176-10177 (2005).
233. Kielec, J. M., Valentine, K. G., Babu, C. R. & Wand, A. J. Reverse micelles in integral membrane protein structural biology by solution NMR spectroscopy. *Structure* **17**, 345-351 (2009).
234. Hwang, P. M. & Kay, L. E. Solution structure and dynamics of integral membrane proteins by NMR: a case study involving the enzyme PagP. *Methods Enzymol* **394**, 335-350 (2005).
235. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J Biomol NMR* **13**, 369-374 (1999).
236. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36**, 1389-1401 (1997).
237. Hilty, C., Fernandez, C., Wider, G. & Wüthrich, K. Side chain NMR assignments in the membrane protein OmpX reconstituted in DHPC micelles. *J Biomol NMR* **23**, 289-301 (2002).
238. Ayala, I., Sounier, R., Use, N., Gans, P. & Boisbouvier, J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR* **43**, 111-119 (2009).
239. Mueller, G. A. et al. Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J Mol Biol* **300**, 197-212 (2000).
240. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114 (1997).
241. Prestegard, J. H. New techniques in structural NMR--anisotropic interactions. *Nat Struct Biol* **5 Suppl**, 517-522 (1998).
242. Gaemers, S. & Bax, A. Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. *J Am Chem Soc* **123**, 12343-12352 (2001).
243. Clore, G. M. & Gronenborn, A. M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc Natl Acad Sci U S A* **95**, 5891-5898 (1998).
244. Williams, D. C. J., Cai, M. & Clore, G. M. Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. *J Biol Chem* **279**, 1449-1457 (2004).
245. Ishii, Y., Markus, M. A. & Tycko, R. Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *J Biomol NMR* **21**, 141-151 (2001).
246. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR* **18**, 303-309 (2000).
247. Douglas, S. M., Chou, J. J. & Shih, W. M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc Natl Acad Sci U S A* **104**, 6644-6648 (2007).
248. Lorieau, J., Yao, L. & Bax, A. Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins. *J Am Chem Soc* **130**, 7536-7537 (2008).
249. Ma, C. & Opella, S. J. Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy. *J Magn Reson* **146**, 381-384 (2000).
250. Chou, J. J., Gaemers, S., Howder, B., Louis, J. M. & Bax, A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR* **21**, 377-382 (2001).
251. Cierpicki, T., Liang, B., Tamm, L. K. & Bushweller, J. H. Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A. *J Am Chem Soc* **128**, 6947-6951 (2006).
252. Bloembergen, N. & Morgan, L. O. *Proton relaxation times in paramagnetic solutions. Effects of*

electron spin relaxation. 1961).
253. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365 (2000).
254. Iwahara, J., Anderson, D. E., Murphy, E. C. & Clore, G. M. EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J Am Chem Soc* **125**, 6634-6635 (2003).
255. Kosen, P. A. Spin labeling of proteins. *Methods Enzymol* **177**, 86-121 (1989).
256. Bertini, I., Donaire, A., Felli, I. C., Rosato, A. & Luchinat, C. *From NOESY cross peaks to structural constraints in a paramagnetic metalloprotein* UNIV BOLOGNA,INST AGR CHEM,BOLOGNA,ITALY, 1996).
257. Bertini, I., Luchinat, C. & Monnanni, R. Interaction of carbon dioxide and copper (II) carbonic anhydrase. *Journal of the American Chemical Society* **109**, 7855 (1987).
258. Bertini, I. & Luchinat, C. Paramagnetic constraints: An aid for quick solution structure determination of paramagnetic metalloproteins. *Concepts in Magnetic Resonance* **14**, 259 (2002).
259. Hubbell, W. L. & Altenbach, C. *Investigation of structure and dynamics in membrane-proteins using site-directed spin labeling* 1994).
260. Huang, J. R. & Grzesiek, S. *Ensemble Calculations of Unstructured Proteins Constrained by RDC and PRE Data: A Case Study of Urea-Denatured Ubiquitin* Univ Basel, Biozentrum, Div Struct Biol, CH-4056 Basel, Switzerland stephan.grzesiek@unibas.ch, 2010).
261. Ganguly, D. & Chen, J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J Mol Biol* **390**, 467-477 (2009).
262. Gillespie, J. R. & Shortle, D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* **268**, 158-169 (1997).
263. Gillespie, J. R. & Shortle, D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J Mol Biol* **268**, 170-184 (1997).
264. Salmon, L. et al. NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* **132**, 8407-8418 (2010).
265. Wu, K. P. & Baum, J. Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement. *J Am Chem Soc* **132**, 5546-5547 (2010).
266. Liang, B., Bushweller, J. H. & Tamm, L. K. Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* **128**, 4389-4397 (2006).
267. Roosild, T. P. et al. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **307**, 1317-1321 (2005).
268. Bermel, W. et al. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl* **44**, 3089-3092 (2005).
269. Bertini, I., Jimenez, B., Pierattelli, R., Wedd, A. G. & Xiao, Z. Protonless 13C direct detection NMR: characterization of the 37 kDa trimeric protein CutA1. *Proteins* **70**, 1196-1205 (2008).
270. Nietlispach, D. Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J Biomol NMR* **31**, 161-166 (2005).
271. Salzmann, M., Wider, G., Pervushin, K. & Wüthrich, K. Improved sensitivity and coherence selection for [15N,1H]-TROSY elements in triple resonance experiments. *J Biomol NMR* **15**, 181-184 (1999).
272. Chung, J. & Kroon, G. 1H,15N,13C-triple resonance NMR of very large systems at 900 MHz. *J Magn Reson* **163**, 360-368 (2003).
273. Hu, K., Eletsky, A. & Pervushin, K. Backbone resonance assignment in large protonated proteins using a combination of new 3D TROSY-HN(CA)HA, 4D TROSY-HACANH and 13C-detected HACACO experiments. *J Biomol NMR* **26**, 69-77 (2003).
274. Loria, J. P., Rance, M. & Palmer, A. G. r. Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J Magn Reson* **141**, 180-184 (1999).
275. Zhu, G., Xia, Y., Lin, D. & Gao, X. TROSY-based correlation and NOE spectroscopy for NMR structural studies of large proteins. *Methods Mol Biol* **278**, 57-78 (2004).
276. Miclet, E. et al. Relaxation-optimized NMR spectroscopy of methylene groups in proteins and nucleic acids. *J Am Chem Soc* **126**, 10560-10570 (2004).
277. Tugarinov, V., Sprangers, R. & Kay, L. E. Line narrowing in methyl-TROSY using zero-

quantum 1H-13C NMR spectroscopy. *J Am Chem Soc* **126**, 4921-4925 (2004).
278. Tugarinov, V., Hwang, P. M. & Kay, L. E. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* **73**, 107-146 (2004).
279. Furukawa, H. et al. Conformation of ligands bound to the muscarinic acetylcholine receptor. *Mol Pharmacol* **62**, 778-787 (2002).
280. Inooka, H. et al. Conformation of a peptide ligand bound to its G-protein coupled receptor. *Nat Struct Biol* **8**, 161-165 (2001).
281. Kisselev, O. G. et al. Light-activated rhodopsin induces structural binding motif in G protein alpha subunit. *Proc Natl Acad Sci U S A* **95**, 4270-4275 (1998).
282. Nisius, L., Rogowski, M., Vangelista, L. & Grzesiek, S. Large-scale expression and purification of the major HIV-1 coreceptor CCR5 and characterization of its interaction with RANTES. *Protein Expr Purif* **61**, 155-162 (2008).
283. Klein-Seetharaman, J. et al. Solution NMR spectroscopy of [alpha -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci U S A* **99**, 3452-3457 (2002).
284. Klein-Seetharaman, J. et al. Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad Sci U S A* **101**, 3409-3413 (2004).
285. Loewen, M. C. et al. Solution 19F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proc Natl Acad Sci U S A* **98**, 4888-4892 (2001).
286. Gautier, A., Kirkpatrick, J. P. & Nietlispach, D. Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed Engl* **47**, 7297-7300 (2008).
287. Reeves, P. J., Thurmond, R. L. & Khorana, H. G. Structure and function in rhodopsin: high level expression of a synthetic bovine opsin gene and its mutants in stable mammalian cell lines. *Proc Natl Acad Sci U S A* **93**, 11487-11492 (1996).
288. De Grip, W. J. Purification of bovine rhodopsin over concanavalin A--sepharose. *Methods Enzymol* **81**, 197-207 (1982).
289. Whitehead, B., Craven, C. J. & Waltho, J. P. Double and triple resonance NMR methods for protein assignment. *Methods Mol Biol* **60**, 29-52 (1997).
290. Tamm, L. K., Abildgaard, F., Arora, A., Blad, H. & Bushweller, J. H. Structure, dynamics and function of the outer membrane protein A (OmpA) and influenza hemagglutinin fusion domain in detergent micelles by solution NMR. *FEBS Lett* **555**, 139-143 (2003).
291. Orekhov, V. Y., Pervushin, K. V. & Arseniev, A. S. Backbone dynamics of (1-71)bacterioopsin studied by two-dimensional 1H-15N NMR spectroscopy. *Eur J Biochem* **219**, 887-896 (1994).
292. Orekhov, V. Y., Abdulaeva, G. V., Musina, L. Y. & Arseniev, A. S. 1H-15N-NMR studies of bacteriorhodopsin Halobacterium halobium. Conformational dynamics of the four-helical bundle. *Eur J Biochem* **210**, 223-229 (1992).
293. Oxenoid, K., Sonnichsen, F. D. & Sanders, C. R. Topology and secondary structure of the N-terminal domain of diacylglycerol kinase. *Biochemistry* **41**, 12876-12882 (2002).
294. Tong, K. I., Yamamoto, M. & Tanaka, T. A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. *J Biomol NMR* **42**, 59-67 (2008).
295. Brüggert, M., Rehm, T., Shanker, S., Georgescu, J. & Holak, T. A. A novel medium for expression of proteins selectively labeled with 15N-amino acids in Spodoptera frugiperda (Sf9) insect cells. *J Biomol NMR* **25**, 335-348 (2003).
296. Wang, Q., Parrish, A. R. & Wang, L. Expanding the genetic code for biological studies. *Chem Biol* **16**, 323-336 (2009).
297. Xie, J. & Schultz, P. G. Adding amino acids to the genetic repertoire. *Curr Opin Chem Biol* **9**, 548-554 (2005).
298. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct* **25**, 163-195 (1996).
299. Klein-Seetharaman, J., Getmanova, E. V., Loewen, M. C., Reeves, P. J. & Khorana, H. G. NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: applicability of solution 19F NMR. *Proc Natl Acad Sci U S A* **96**, 13744-13749 (1999).
300. Raport, C. J., Gosling, J., Schweickart, V. L., Gray, P. W. & Charo, I. F. Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1beta, and MIP-1alpha. *J Biol Chem* **271**, 17161-17166 (1996).
301. Sodhi, A., Montaner, S. & Gutkind, J. S. Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat Rev Mol Cell Biol* **5**, 998-1012 (2004).
302. Paterlini, M. G. Structure modeling of the chemokine receptor CCR5: implications for ligand

binding and selectivity. *Biophys J* **83**, 3012-3031 (2002).
303. Huang, C. C. et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317**, 1930-1934 (2007).
304. Duma, L., Haussinger, D., Rogowski, M., Lusso, P. & Grzesiek, S. Recognition of RANTES by extracellular parts of the CCR5 receptor. *J Mol Biol* **365**, 1063-1075 (2007).
305. Baldwin, J. M. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J* **12**, 1693-1703 (1993).
306. Liu, J., Conklin, B. R., Blin, N., Yun, J. & Wess, J. Identification of a receptor/G-protein contact site critical for signaling specificity and G-protein activation. *Proc Natl Acad Sci U S A* **92**, 11642-11646 (1995).
307. Taylor, J. M., Jacob-Mosier, G. G., Lawton, R. G., Remmers, A. E. & Neubig, R. R. Binding of an alpha 2 adrenergic receptor third intracellular loop peptide to G beta and the amino terminus of G alpha. *J Biol Chem* **269**, 27618-27624 (1994).
308. Taylor, J. M., Jacob-Mosier, G. G., Lawton, R. G., VanDort, M. & Neubig, R. R. Receptor and membrane interaction sites on Gbeta. A receptor-derived peptide binds to the carboxyl terminus. *J Biol Chem* **271**, 3336-3339 (1996).
309. Bourne, H. R. How receptors talk to trimeric G proteins. *Curr Opin Cell Biol* **9**, 134-142 (1997).
310. Hamm, H. E. & Gilchrist, A. Heterotrimeric G proteins. *Curr Opin Cell Biol* **8**, 189-196 (1996).
311. Ratnala, V. R. & Kobilka, B. Understanding the ligand-receptor-G protein ternary complex for GPCR drug discovery. *Methods Mol Biol* **552**, 67-77 (2009).
312. West, R. E. J., Moss, J., Vaughan, M., Liu, T. & Liu, T. Y. Pertussis toxin-catalyzed ADP-ribosylation of transducin. Cysteine 347 is the ADP-ribose acceptor site. *J Biol Chem* **260**, 14428-14430 (1985).
313. Clore, G. M. & Gronenborn, A. M. Theory and applications of the transferred nuclear overhauser effect to the study of the conformations of small ligands bound to proteins. *J. Magn. Reson* **48**, 402-417 (1982).
314. Behling, R. W., Yamane, T., Navon, G. & Jelinski, L. W. Conformation of acetylcholine bound to the nicotinic acetylcholine receptor. *Proc Natl Acad Sci U S A* **85**, 6721-6725 (1988).
315. Portoghese, P. S. Relationships between stereostructure and pharmacological activities. *Annu Rev Pharmacol* **10**, 51-76 (1970).
316. Colmenares, L. U., Niemczura, W. P., Asato, A. E. & Liu, R. S. H. A 19F NMR Study of Rhodopsin Analogs: Use of Vinylfluororetinal Chromophores. *J. Phys. Chem* **100**, 9175-9180 (1996).
317. Fujimoto, Y. et al. On the bioactive conformation of the rhodopsin chromophore: absolute sense of twist around the 6-s-cis bond. *Chemistry* **7**, 4198-4204 (2001).
318. Salgado, G. F. et al. Deuterium NMR structure of retinal in the ground state of rhodopsin. *Biochemistry* **43**, 12819-12828 (2004).
319. Smith, S. O., Courtin, J., de Groot, H., Gebhard, R. & Lugtenburg, J. 13C magic-angle spinning NMR studies of bathorhodopsin, the primary photoproduct of rhodopsin. *Biochemistry* **30**, 7409-7415 (1991).
320. Smith, S. O. et al. Low-temperature solid-state 13C NMR studies of the retinal chromophore in rhodopsin. *Biochemistry* **26**, 1606-1611 (1987).
321. Spooner, P. J. et al. Relative orientation between the beta-ionone ring and the polyene chain for the chromophore of rhodopsin in native membranes. *Biochemistry* **41**, 7549-7555 (2002).
322. Ahuja, S. et al. Location of the retinal chromophore in the activated state of rhodopsin*. *J Biol Chem* **284**, 10190-10201 (2009).
323. Ahuja, S. et al. 6-s-cis Conformation and polar binding pocket of the retinal chromophore in the photoactivated state of rhodopsin. *J Am Chem Soc* **131**, 15160-15169 (2009).
324. Ahuja, S. et al. Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nat Struct Mol Biol* **16**, 168-175 (2009).
325. Spooner, P. J. et al. Conformational similarities in the beta-ionone ring region of the rhodopsin chromophore in its ground state and after photoactivation to the metarhodopsin-I intermediate. *Biochemistry* **42**, 13371-13378 (2003).
326. Colmenares, L. U. et al. NMR studies of fluorinated visual pigment analogs. *Biochem Biophys Res Commun* **179**, 1337-1343 (1991).
327. Luca, S. et al. The conformation of neurotensin bound to its G protein-coupled receptor. *Proc Natl Acad Sci U S A* **100**, 10706-10711 (2003).
328. Ratnala, V. R. et al. Solid-state NMR evidence for a protonation switch in the binding pocket of the H1 receptor upon binding of the agonist histamine. *J Am Chem Soc* **129**, 867-872 (2007).
329. Hoff, W. D., Jung, K. H. & Spudich, J. L. Molecular mechanism of photosignaling by archaeal

sensory rhodopsins. *Annu Rev Biophys Biomol Struct* **26**, 223-258 (1997).
330. Barna, J. C. J. & Laue, E. D. Conventional and exponential sampling for 2 D NMR experiments with. *Journal of magnetic resonance* (1987).
331. Rovnyak, D. et al. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *Journal of Magnetic Resonance* **170**, 15 (2004).
332. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc* **121**, 844 (1999).
333. Metzler, W. J. et al. Characterization of the three-dimensional solution structure of human profilin: 1H, 13C, and 15N NMR assignments and global folding pattern. *Biochemistry* **32**, 13818-13829 (1993).
334. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N. & Spudich, J. L. Crystal structure of sensory rhodopsin II at 2.4 angstroms: insights into color tuning and transducer interaction. *Science* **293**, 1499-1503 (2001).
335. Etzkorn, M. et al. Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew Chem Int Ed Engl* **46**, 459-462 (2007).
336. Tian, C. et al. Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor. *J Am Chem Soc* **127**, 8010-8011 (2005).
337. Tian, C. et al. Solution NMR Spectroscopy of the Human Vasopressin V2 Receptor, A G Protein-Coupled Receptor. *J Am Chem Soc* **128**, 5300 (2006).
338. Werner, K., Richter, C., Klein-Seetharaman, J. & Schwalbe, H. Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* **40**, 49-53 (2008).
339. Andrew, E. R. & Bradbury, A. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* **182**, 1659 (1958).
340. Davis, J. H. Static and magic angle spinning NMR of membrane peptides and proteins. *Progress in nuclear magnetic resonance spectroscopy* **35**, 1 (1999).
341. Cross, T. A. Solid-state NMR structural studies of peptides and proteins in membranes. *Current Opinion in Structural Biology* **574**, 4 (1994).
342. Opella, S. J. & Marassi, F. M. Structure determination of membrane proteins by NMR spectroscopy. *Chem Rev* **104**, 3587-3606 (2004).
343. Werner, K. et al. Combined solid state and solution NMR studies of alpha,epsilon-15N labeled bovine rhodopsin. *J Biomol NMR* **37**, 303-312 (2007).
344. Grobner, G. et al. Observations of light-induced structural changes of retinal within rhodopsin. *Nature* **405**, 810-813 (2000).
345. Schmidt, P. et al. A reconstitution protocol for the in vitro folded human G protein-coupled Y2 receptor into lipid environment. *Biophys Chem* **150**, 29-36 (2010).
346. Tapaneeyakorn, S., Goddard, A. D., Oates, J., Willis, C. L. & Watts, A. Solution- and solid-state NMR studies of GPCRs and their ligands. *Biochim Biophys Acta* (2010).
347. Katragadda, M. et al. Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin. *J Pept Res* **58**, 79-89 (2001).
348. Callihan, D. E. & Logan, T. M. Conformations of peptide fragments from the FK506 binding protein: comparison with the native and urea-unfolded states. *J Mol Biol* **285**, 2161-2175 (1999).
349. Hunt, J. F. et al. A biophysical study of integral membrane protein folding. *Biochemistry* **36**, 15156-15176 (1997).
350. Yang, A. S., Hitz, B. & Honig, B. Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding. *J Mol Biol* **259**, 873-882 (1996).
351. Marti, T. Refolding of bacteriorhodopsin from expressed polypeptide fragments. *J Biol Chem* **273**, 9312-9322 (1998).
352. Martin, N. P., Leavitt, L. M., Sommers, C. M. & Dumont, M. E. Assembly of G protein-coupled receptors from fragments: identification of functional receptors with discontinuities in each of the loops connecting transmembrane segments. *Biochemistry* **38**, 682-695 (1999).
353. Konig, B. et al. Three cytoplasmic loops of rhodopsin interact with transducin. *Proc Natl Acad Sci U S A* **86**, 6878-6882 (1989).
354. Yeagle, P. L., Alderfer, J. L. & Albert, A. D. Structure of the carboxy-terminal domain of bovine rhodopsin. *Nat Struct Biol* **2(10)**, 832-834 (1995).
355. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S. & Sykes, B. D. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry*

31, 8790-8798 (1992).
356. Franzoni, L. et al. Structure of the C-terminal fragment 300-320 of the rat angiotensin II AT1A receptor and its relevance with respect to G-protein coupling. *J Biol Chem* **272**, 9734-9741 (1997).
357. Nicastro, G. et al. Conformational features of a synthetic model of the first extracellular loop of the angiotensin II AT1A receptor. *J Pept Sci* **9**, 229-243 (2003).
358. Salinas, R. K. et al. Trifluoroethanol and binding to model membranes stabilize a predicted turn in a peptide corresponding to the first extracellular loop of the angiotensin II AT(1A) receptor. *Biopolymers* **65**, 21-31 (2002).
359. Mierke, D. F., Royo, M., Pellegrini, M., Sun, H. & Chorev, M. Peptide Mimetic of the Third Cytoplasmic Loop of the PTH/PTHrP Receptor. *J Am Chem Soc* **118**, 8998-9004 (1996).
360. Pellegrini, M., Royo, M., Chorev, M. & Mierke, D. F. Conformational characterization of a peptide mimetic of the third cytoplasmic loop of the G-protein coupled parathyroid hormone/parathyroid hormone related protein receptor. *Biopolymers* **40**, 653-666 (1996).
361. Piserchio, A., Bisello, A., Rosenblatt, M., Chorev, M. & Mierke, D. F. Characterization of parathyroid hormone/receptor interactions: structure of the first extracellular loop. *Biochemistry* **39**, 8153-8160 (2000).
362. Ulfers, A. L., McMurry, J. L., Kendall, D. A. & Mierke, D. F. Structure of the third intracellular loop of the human cannabinoid 1 receptor. *Biochemistry* **41**, 11344-11350 (2002).
363. Choi, G., Guo, J. & Makriyannis, A. The conformation of the cytoplasmic helix 8 of the CB1 cannabinoid receptor using NMR and circular dichroism. *Biochim Biophys Acta* **1668**, 1-9 (2005).
364. Zheng, H., Zhao, J., Sheng, W. & Xie, X. Q. A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization. *Biopolymers* **83**, 46-61 (2006).
365. Zhao, J., Zheng, H. & Xie, X. Q. NMR characterization of recombinant transmembrane protein CB2 fragment CB2(180-233). *Protein Pept Lett* **13**, 335-342 (2006).
366. Giragossian, C. & Mierke, D. F. Intermolecular interactions between cholecystokinin-8 and the third extracellular loop of the cholecystokinin A receptor. *Biochemistry* **40**, 3804-3809 (2001).
367. Pellegrini, M. & Mierke, D. F. Structural characterization of peptide hormone/receptor interactions by NMR spectroscopy. *Biopolymers* **51**, 208-220 (1999).
368. Giragossian, C. & Mierke, D. F. Intermolecular interactions between cholecystokinin-8 and the third extracellular loop of the cholecystokinin-2 receptor. *Biochemistry* **41**, 4560-4566 (2002).
369. Arshava, B., Taran, I., Xie, H., Becker, J. M. & Naider, F. High resolution NMR analysis of the seven transmembrane domains of a heptahelical receptor in organic-aqueous medium. *Biopolymers* **64**, 161-176 (2002).
370. Naider, F. et al. Synthetic peptides as probes for conformational preferences of domains of membrane receptors. *Biopolymers* **80**, 199-213 (2005).
371. Estephan, R. et al. Biosynthesis and NMR analysis of a 73-residue domain of a Saccharomyces cerevisiae G protein-coupled receptor. *Biochemistry* **44**, 11795-11810 (2005).
372. Neumoin, A. et al. Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles. *Biophys J* **96**, 3187-3196 (2009).
373. Yeagle, P. L. et al. Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin. *J Pept Res* **55**, 455-465 (2000).
374. Chopra, A., Yeagle, P. L., Alderfer, J. A. & Albert, A. D. Solution structure of the sixth transmembrane helix of the G-protein-coupled receptor, rhodopsin. *Biochim Biophys Acta* **1463**, 1-5 (2000).
375. Yeagle, P. L., Alderfer, J. L., Salloum, A. C., Ali, L. & Albert, A. D. The first and second cytoplasmic loops of the G-protein receptor, rhodopsin, independently form beta-turns. *Biochemistry* **36**, 3864-3869 (1997).
376. Yeagle, P. L., Alderfer, J. L. & Albert, A. D. Structure of the third cytoplasmic loop of bovine rhodopsin. *Biochemistry* **34**, 14621-14625 (1995).
377. Krishna, A. G., Menon, S. T., Terry, T. J. & Sakmar, T. P. Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch. *Biochemistry* **41**, 8298-8309 (2002).
378. Piserchio, A. et al. Structural insight into the role of the second intracellular loop of the bradykinin 2 receptor in signaling and internalization. *Biopolymers* **63**, 239-246 (2002).
379. Piserchio, A. et al. Bradykinin B2 receptor signaling: structural and functional characterization of the C-terminus. *Biopolymers* **80**, 367-373 (2005).
380. Ulfers, A. L., Piserchio, A. & Mierke, D. F. Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P. *Biopolymers* **66**, 339-349 (2002).

381. Macdonald, D. et al. Photoaffinity labeling of mutant neurokinin-1 receptors reveals additional structural features of the substance P/NK-1 receptor complex. *Biochemistry* **40**, 2530-2539 (2001).
382. Pellegrini, M., Bremer, A. A., Ulfers, A. L., Boyd, N. D. & Mierke, D. F. Molecular characterization of the substance P*neurokinin-1 receptor complex: development of an experimentally based model. *J Biol Chem* **276**, 22862-22867 (2001).
383. Jung, H., Windhaber, R., Palm, D. & Schnackerz, K. D. NMR and circular dichroism studies of synthetic peptides derived from the third intracellular loop of the beta-adrenoceptor. *FEBS Lett* **358**, 133-136 (1995).
384. Katragadda, M., Maciejewski, M. W. & Yeagle, P. L. Structural studies of the putative helix 8 in the human beta(2) adrenergic receptor: an NMR study. *Biochim Biophys Acta* **1663**, 74-81 (2004).
385. Chung, D. A. et al. NMR structure of the second intracellular loop of the alpha 2A adrenergic receptor: evidence for a novel cytoplasmic helix. *Biochemistry* **41**, 3596-3604 (2002).
386. Lazarova, T., Brewin, K. A., Stoeber, K. & Robinson, C. R. Characterization of peptides corresponding to the seven transmembrane domains of human adenosine A2a receptor. *Biochemistry* **43**, 12945-12954 (2004).
387. Demene, H. et al. Active peptidic mimics of the second intracellular loop of the V(1A) vasopressin receptor are structurally related to the second intracellular rhodopsin loop: a combined 1H NMR and biochemical study. *Biochemistry* **42**, 8204-8213 (2003).
388. Ruan, K. H. et al. Solution structure of the second extracellular loop of human thromboxane A2 receptor. *Biochemistry* **40**, 275-280 (2001).
389. Wu, J., So, S. P. & Ruan, K. H. Solution structure of the third extracellular loop of human thromboxane A2 receptor. *Arch Biochem Biophys* **414**, 287-293 (2003).
390. Kimmel, J. R., Hayden, L. J. & Pollock, H. G. Isolation and characterization of a new pancreatic polypeptide hormone. *J Biol Chem* **250**, 9369-9376 (1975).
391. Hazelwood, R. L. The pancreatic polypeptide (PP-fold) family: gastrointestinal, vascular, and feeding behavioral implications. *Proc Soc Exp Biol Med* **202**, 44-63 (1993).
392. Trinh, T., van Dumont, Y. & Quirion, R. High levels of specific neuropeptide Y/pancreatic polypeptide receptors in the rat hypothalamus and brainstem. *Eur J Pharmacol* **318**, R1-3 (1996).
393. Clark, J. T., Kalra, P. S., Crowley, W. R. & Kalra, S. P. Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427-429 (1984).
394. Inui, A. et al. Neuropeptide regulation of feeding in dogs. *Am J Physiol* **261**, R588-94 (1991).
395. Tatemoto, K. & Mutt, V. Chemical determination of polypeptide hormones. *Proc Natl Acad Sci U S A* **75**, 4115-4119 (1978).
396. Tatemoto, K. Isolation and characterization of peptide YY (PYY), a candidate gut hormone that inhibits pancreatic exocrine secretion. *Proc Natl Acad Sci U S A* **79**, 2514-2518 (1982).
397. Tatemoto, K., Carlquist, M. & Mutt, V. Neuropeptide Y--a novel brain peptide with structural similarities to peptide YY and pancreatic polypeptide. *Nature* **296**, 659-660 (1982).
398. Tatemoto, K. & Mutt, V. Isolation of two novel candidate hormones using a chemical method for finding naturally occurring polypeptides. *Nature* **285**, 417-418 (1980).
399. Nakajima, M. et al. Effects of pancreatic polypeptide family peptides on feeding and learning behavior in mice. *J Pharmacol Exp Ther* **268**, 1010-1014 (1994).
400. Lopez-Valpuesta, F. J., Nyce, J. W., Griffin-Biggs, T. A., Ice, J. C. & Myers, R. D. Antisense to NPY-Y1 demonstrates that Y1 receptors in the hypothalamus underlie NPY hypothermia and feeding in rats. *Proc Biol Sci* **263**, 881-886 (1996).
401. Erickson, J. C., Clegg, K. E. & Palmiter, R. D. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415-421 (1996).
402. Naveilhan, P., Canals, J. M., Arenas, E. & Ernfors, P. Distinct roles of the Y1 and Y2 receptors on neuropeptide Y-induced sensitization to sedation. *J Neurochem* **78**, 1201-1207 (2001).
403. Michel, M. C. et al. XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors. *Pharmacol Rev* **50**, 143-150 (1998).
404. Heilig, M. & Widerlov, E. Neurobiology and clinical aspects of neuropeptide Y. *Crit Rev Neurobiol* **9**, 115-136 (1995).
405. Franco-Cereceda, A. & Liska, J. Neuropeptide Y Y1 receptors in vascular pharmacology. *Eur J Pharmacol* **349**, 1-14 (1998).
406. Morton, K. D., McCloskey, M. J. & Potter, E. K. Cardiorespiratory responses to intracerebroventricular injection of neuropeptide Y in anaesthetised dogs. *Regul Pept* **81**, 81-88

(1999).
407. Gehlert, D. R. Multiple receptors for the pancreatic polypeptide (PP-fold) family: physiological implications. *Proc Soc Exp Biol Med* **218**, 7-22 (1998).
408. Dotsch, J., Christiansen, H., Hanze, J., Lampert, F. & Rascher, W. Plasma neuropeptide Y of children with neuroblastoma in relation to stage, age and prognosis, and tissue neuropeptide Y. *Regul Pept* **75-76**, 185-190 (1998).
409. Carr, L. G. et al. A quantitative trait locus for alcohol consumption in selectively bred rat lines. *Alcohol Clin Exp Res* **22**, 884-887 (1998).
410. Thiele, T. E., Marsh, D. J., Ste Marie, L., Bernstein, I. L. & Palmiter, R. D. Ethanol consumption and resistance are inversely related to neuropeptide Y levels. *Nature* **396**, 366-369 (1998).
411. Ilveskoski, E. et al. Association of neuropeptide y polymorphism with the occurrence of type 1 and type 2 alcoholism. *Alcohol Clin Exp Res* **25**, 1420-1422 (2001).
412. Bagnasco, M., Kalra, P. S. & Kalra, S. P. Ghrelin and leptin pulse discharge in fed and fasted rats. *Endocrinology* **143**, 726-729 (2002).
413. Kojima, M. et al. Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656-660 (1999).
414. Kojima, M., Hosoda, H., Matsuo, H. & Kangawa, K. Ghrelin: discovery of the natural endogenous ligand for the growth hormone secretagogue receptor. *Trends Endocrinol Metab* **12**, 118-122 (2001).
415. Stephens, T. W. et al. The role of neuropeptide Y in the antiobesity action of the obese gene product. *Nature* **377**, 530-532 (1995).
416. Gray, T. S. & Morley, J. E. Neuropeptide Y: anatomical distribution and possible function in mammalian nervous system. *Life Sci* **38**, 389-401 (1986).
417. Dumont, Y., Martel, J. C., Fournier, A., St-Pierre, S. & Quirion, R. Neuropeptide Y and neuropeptide Y receptor subtypes in brain and peripheral tissues. *Prog Neurobiol* **38**, 125-167 (1992).
418. Lundberg, J. M. Pharmacology of cotransmission in the autonomic nervous system: integrative aspects on amines, neuropeptides, adenosine triphosphate, amino acids and nitric oxide. *Pharmacol Rev* **48**, 113-178 (1996).
419. Sundler F, B. G., Eckblad E, Hakanson R. *The Biology of Neuropeptide Y and Related Peptides* (Humana Press, Totowa, NJ, 1993).
420. Schwartz, T. W. *Molecular biology of Islet of Langerhans* (Cambridge University Press, Cambridge, 1990).
421. Barr, P. J. Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell* **66**, 1-3 (1991).
422. Eipper, B. A., Mains, R. E. & Glembotski, C. C. Identification in pituitary tissue of a peptide alpha-amidation activity that acts on glycine-extended peptides and requires molecular oxygen, copper, and ascorbic acid. *Proc Natl Acad Sci U S A* **80**, 5144-5148 (1983).
423. Eipper, B. A. et al. Structure of the precursor to an enzyme mediating COOH-terminal amidation in peptide biosynthesis. *Mol Endocrinol* **1**, 777-790 (1987).
424. Fricker, L. D., Evans, C. J., Esch, F. S. & Herbert, E. Cloning and sequence analysis of cDNA for bovine carboxypeptidase E. *Nature* **323**, 461-464 (1986).
425. Larhammar, D. Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11 (1996).
426. Conlon, J. M. The origin and evolution of peptide YY (PYY) and pancreatic polypeptide (PP). *Peptides* **23**, 269-278 (2002).
427. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A* **78**, 4175-4179 (1981).
428. Li, X. A., Sutcliffe, M. J., Schwartz, T. W. & Dobson, C. M. Sequence-specific 1H NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **31**, 1245-1253 (1992).
429. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R. J. Solution structure of monomeric peptide YY supports the functional significance of the PP-fold. *Biochemistry* **39**, 9935-9942 (2000).
430. Lerch, M., Mayrhofer, M. & Zerbe, O. Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J. Mol. Biol.* **339**, 1153-1168 (2004).
431. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. Solution structure of human

neuropeptide Y. *J Biomol NMR* **8**, 379-390 (1996).
432. Cowley, D. J., Hoflack, J. M., Pelton, J. T. & Saudek, V. Structure of neuropeptide Y dimer in solution. *Eur J Biochem* **205**, 1099-1106 (1992).
433. Nordmann, A., Blommers, M. J., Fretz, H., Arvinte, T. & Drake, A. F. Aspects of the molecular structure and dynamics of neuropeptide Y. *Eur J Biochem* **261**, 216-226 (1999).
434. Bettio, A., Dinger, M. C. & Beck-Sickinger, A. G. The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci* **11**, 1834-1844 (2002).
435. Chang, P. J., Noelken, M. E. & Kimmel, J. R. Reversible dimerization of avian pancreatic polypeptide. *Biochemistry* **19**, 1844-1849 (1980).
436. Zimmerberg, J. & Kozlov, M. M. How proteins produce cellular membrane curvature. *Nat Rev Mol Cell Biol* **7**, 9-19 (2006).
437. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. Structure and dynamics of micelle-bound neuropeptide Y: comparison with unligated NPY and implications for receptor selection. *J Mol Biol* **305**, 307-329 (2001).
438. Lerch, M. et al. Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J. Mol. Biol.* **322**, 1117-1133 (2002).
439. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3**, 842-848 (1996).
440. Pellegrini, M., Royo, M., Rosenblatt, M., Chorev, M. & Mierke, D. F. Addressing the tertiary structure of human parathyroid hormone-(1-34). *J Biol Chem* **273**, 10420-10427 (1998).
441. Sargent, D. F. & Schwyzer, R. Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc Natl Acad Sci U S A* **83**, 5774-5778 (1986).
442. Schwyzer, R. Molecular mechanism of opioid receptor selection. *Biochemistry* **25**, 6335-6342 (1986).
443. Schwyzer, R. Peptide-membrane interactions and a new principle in quantitative structure-activity relationships. *Biopolymers* **31**, 785-792 (1991).
444. Schwyzer, R. 100 years lock-and-key concept: are peptide keys shaped and guided to their receptors by the target cell membrane? *Biopolymers* **37**, 5-16 (1995).
445. Schwyzer, R. In search of the 'bio-active conformation'--is it induced by the target cell membrane? *J Mol Recognit* **8**, 3-8 (1995).
446. Bader, R. & Zerbe, O. Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state? *ChemBioChem* **6**, 1520-1534 (2005).
447. Beck-Sickinger, A. G. et al. Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur. J. Biochem.* **225**, 947-958 (1994).
448. Dougherty, D. A. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163-168 (1996).
449. Cabrele, C., Wieland, H. A., Koglin, N., Stidsen, C. & Beck-Sickinger, A. G. Ala31-Aib32: identification of the key motif for high affinity and selectivity of neuropeptide Y at the Y5-receptor. *Biochemistry* **41**, 8043-8049 (2002).
450. Bader, R., Rytz, G., Lerch, M., Beck-Sickinger, A. G. & Zerbe, O. Key motif to gain selectivity at the neuropeptide Y5-receptor: structure and dynamics of micelle-bound [Ala31, Pro32]-NPY. *Biochemistry* **41**, 8031-8042 (2002).
451. Lerch, M. et al. Strongly altered receptor binding properties in PP and NPY chimeras are accompanied by changes in structure and membrane binding. *Biochemistry* **44**, 9255-9264 (2005).
452. Cabrele, C., Wieland, H. A., Langer, M., Stidsen, C. E. & Beck-Sickinger, A. G. Y-receptor affinity modulation by the design of pancreatic polypeptide/neuropeptide Y chimera led to Y(5)-receptor ligands with picomolar affinity. *Peptides* **22**, 365-378 (2001).
453. Herzog, H. et al. Cloned human neuropeptide Y receptor couples to two different second messenger systems. *Proc Natl Acad Sci U S A* **89**, 5794-5798 (1992).
454. Krause, J., Eva, C., Seeburg, P. H. & Sprengel, R. Neuropeptide Y1 subtype pharmacology of a recombinantly expressed neuropeptide receptor. *Mol Pharmacol* **41**, 817-821 (1992).
455. Larhammar, D. et al. Cloning and functional expression of a human neuropeptide Y/peptide YY receptor of the Y1 type. *J Biol Chem* **267**, 10935-10938 (1992).
456. Gerald, C. et al. Expression cloning and pharmacological characterization of a human hippocampal neuropeptide Y/peptide YY Y2 receptor subtype. *J Biol Chem* **270**, 26758-26761 (1995).
457. Ingenhoven, N., Eckard, C. P., Gehlert, D. R. & Beck-Sickinger, A. G. Molecular characterization of the human neuropeptide Y Y2-receptor. *Biochemistry* **38**, 6897-6902 (1999).

458. Rose, P. M. et al. Cloning and functional expression of a cDNA encoding a human type 2 neuropeptide Y receptor. *J Biol Chem* **270**, 22661-22664 (1995).
459. Gehlert, D. R. et al. Expression cloning of a human brain neuropeptide Y Y2 receptor. *Mol Pharmacol* **49**, 224-228 (1996).
460. Bard, J. A., Walker, M. W., Branchek, T. A. & Weinshank, R. L. Cloning and functional expression of a human Y4 subtype receptor for pancreatic polypeptide, neuropeptide Y, and peptide YY. *J Biol Chem* **270**, 26762-26765 (1995).
461. Lundell, I. et al. Cloning of a human receptor of the NPY receptor family with high affinity for pancreatic polypeptide and peptide YY. *J Biol Chem* **270**, 29123-29128 (1995).
462. Gerald, C. et al. A receptor subtype involved in neuropeptide-Y-induced food intake. *Nature* **382**, 168-171 (1996).
463. Lee, C. C. & Miller, R. J. Is there really an NPY Y3 receptor? *Regul Pept* **75-76**, 71-78 (1998).
464. Weinberg, D. H. et al. Cloning and expression of a novel neuropeptide Y receptor. *J Biol Chem* **271**, 16435-16438 (1996).
465. Blomqvist, A. G. & Herzog, H. Y-receptor subtypes--how many more? *Trends Neurosci* **20**, 294-298 (1997).
466. Larhammar, D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **65**, 165-174 (1996).
467. Robin-Jagerschmidt, C. et al. The ligand binding site of NPY at the rat Y1 receptor investigated by site-directed mutagenesis and molecular modeling. *Mol Cell Endocrinol* **139**, 187-198 (1998).
468. Wieland, H. A., Eckard, C. P., Doods, H. N. & Beck-Sickinger, A. G. Probing of the neuropeptide Y-Y1-receptors interaction with anti-receptor antibodies. *Eur J Biochem* **255**, 595-603 (1998).
469. Grundemar, L. & Hogestatt, E. D. Unmasking the vasoconstrictor response to neuropeptide Y and its interaction with vasodilating agents in vitro. *Eur J Pharmacol* **221**, 71-76 (1992).
470. Wahlestedt, C. & Hakanson, R. Effects of neuropeptide Y (NPY) at the sympathetic neuroeffector junction. Can pre- and postjunctional receptors be distinguished? *Med Biol* **64**, 85-88 (1986).
471. Wahlestedt, C., Pich, E. M., Koob, G. F., Yee, F. & Heilig, M. Modulation of anxiety and neuropeptide Y-Y1 receptors by antisense oligodeoxynucleotides. *Science* **259**, 528-531 (1993).
472. Potter, E. K. et al. Pre- and postjunctional actions of neuropeptide Y and related peptides. *Regul Pept* **25**, 167-177 (1989).
473. Flood, J. F. & Morley, J. E. Dissociation of the effects of neuropeptide Y on feeding and memory: evidence for pre- and postsynaptic mediation. *Peptides* **10**, 963-966 (1989).
474. Norenberg, W., Bek, M., Limberger, N., Takeda, K. & Illes, P. Inhibition of nicotinic acetylcholine receptor channels in bovine adrenal chromaffin cells by Y3-type neuropeptide Y receptors via the adenylate cyclase/protein kinase A system. *Naunyn Schmiedebergs Arch Pharmacol* **351**, 337-347 (1995).
475. Grundemar, L., Wahlestedt, C. & Reis, D. J. Neuropeptide Y acts at an atypical receptor to evoke cardiovascular depression and to inhibit glutamate responsiveness in the brainstem. *J Pharmacol Exp Ther* **258**, 633-638 (1991).
476. Grundemar, L., Wahlestedt, C. & Reis, D. J. Long-lasting inhibition of the cardiovascular responses to glutamate and the baroreceptor reflex elicited by neuropeptide Y injected into the nucleus tractus solitarius of the rat. *Neurosci Lett* **122**, 135-139 (1991).
477. Schwartz, T. W. Pancreatic polypeptide: a hormone under vagal control. *Gastroenterology* **85**, 1411-1425 (1983).
478. Gehlert, D. R. et al. Characterization of the peptide binding requirements for the cloned human pancreatic polypeptide-preferring receptor. *Mol Pharmacol* **50**, 112-118 (1996).
479. Hu, Y. et al. Identification of a novel hypothalamic neuropeptide Y receptor associated with feeding behavior. *J Biol Chem* **271**, 26315-26319 (1996).
480. Starback, P., Wraith, A., Eriksson, H. & Larhammar, D. Neuropeptide Y receptor gene y6: multiple deaths or resurrections? *Biochem Biophys Res Commun* **277**, 264-269 (2000).
481. Gregor, P., Feng, Y., DeCarr, L. B., Cornfield, L. J. & McCaleb, M. L. Molecular characterization of a second mouse pancreatic polypeptide receptor and its inactivated human homologue. *J Biol Chem* **271**, 27776-27781 (1996).
482. Beck-Sickinger, A. G. *Methods in Molecular Biology, Neuropeptide Protocols* (Humana Press, Totowa, NJ, 1997).
483. Liao, S. et al. De novo design, synthesis, and biological activities of high-affinity and selective non-peptide agonists of the delta-opioid receptor. *J Med Chem* **41**, 4767-4776 (1998).

484. Larhammar, D. & Salaneck, E. Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-151 (2004).
485. Merten, N. et al. Receptor subtype-specific docking of Asp6.59 with C-terminal arginine residues in Y receptor ligands. *J Biol Chem* **282**, 7543-7551 (2007).
486. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J. Biol. Chem.* **269**, 2863-2869 (1994).
487. Sautel, M. et al. Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. *Mol. Pharmacol.* **50**, 285-292 (1996).
488. Walker, P. et al. High level expression of human neuropeptide Y receptors in mammalian cells infected with a recombinant vaccinia virus. *Mol Cell Endocrinol* **91**, 107-112 (1993).
489. Kirby, D. A., Boublik, J. H. & Rivier, J. E. Neuropeptide Y: Y1 and Y2 affinities of the complete series of analogues with single D-residue substitutions. *J Med Chem* **36**, 3802-3808 (1993).
490. Leban, J. J. et al. Novel modified carboxy terminal fragments of neuropeptide Y with high affinity for Y2-type receptors and potent functional antagonism at a Y1-type receptor. *J Med Chem* **38**, 1150-1157 (1995).
491. Rist, B. et al. The bioactive conformation of neuropeptide Y analogues at the human Y2-receptor. *Eur J Biochem* **247**, 1019-1028 (1997).
492. McCrea, K. et al. 2-36[K4,RYYSA(19-23)]PP a novel Y5-receptor preferring ligand with strong stimulatory effect on food intake. *Regul Pept* **87**, 47-58 (2000).
493. Eckard, C. P., Cabrele, C., Wieland, H. A. & Beck-Sickinger, A. Characterisation of Neuropeptide Y Receptor Subtypes by Synthetic NPY Analogues and by Anti-receptor Antibodies. *Molecules* **6**, 448-467 (2001).
494. Criscione, L. et al. Food intake in free-feeding and energy-deprived lean rats is mediated by the neuropeptide Y5 receptor. *J Clin Invest* **102**, 2136-2145 (1998).
495. Haynes, A. C., Arch, J. R., Wilson, S., McClue, S. & Buckingham, R. E. Characterisation of the neuropeptide Y receptor that mediates feeding in the rat: a role for the Y5 receptor? *Regul Pept* **75-76**, 355-361 (1998).
496. Kanatani, A. et al. NPY-induced feeding involves the action of a Y1-like receptor in rodents. *Regul Pept* **75-76**, 409-415 (1998).
497. Marsh, D. J., Hollopeter, G., Kafer, K. E. & Palmiter, R. D. Role of the Y5 neuropeptide Y receptor in feeding and obesity. *Nat Med* **4**, 718-721 (1998).
498. Tang-Christensen, M., Kristensen, P., Stidsen, C. E., Brand, C. L. & Larsen, P. J. Central administration of Y5 receptor antisense decreases spontaneous food intake and attenuates feeding in response to exogenous neuropeptide Y. *J Endocrinol* **159**, 307-312 (1998).
499. Berglund, M. M. et al. Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands. *Peptides* **22**, 351-356 (2001).
500. Beck-Sickinger, A. G. & Jung, G. Structure-activity relationships of neuropeptide Y analogues with respect to Y1 and Y2 receptors. *Biopolymers* **37**, 123-142 (1995).
501. Eriksson, H. et al. The cloned guinea pig pancreatic polypeptide receptor Y4 resembles more the human Y4 than does the rat Y4. *Regul Pept* **75-76**, 29-37 (1998).
502. Eckard, C. P. & Beck-Sickinger, A. G. Characterisation of G-protein-coupled receptors by antibodies. *Curr Med Chem* **7**, 897-910 (2000).
503. Kage, R., Leeman, S. E., Krause, J. E., Costello, C. E. & Boyd, N. D. Identification of methionine as the site of covalent attachment of a p-benzoyl-phenylalanine-containing analogue of substance P on the substance P (NK-1) receptor. *J Biol Chem* **271**, 25797-25800 (1996).
504. Li, S., Liu, X., Min, L. & Ascoli, M. Mutations of the second extracellular loop of the human lutropin receptor emphasize the importance of receptor activation and de-emphasize the importance of receptor phosphorylation in agonist-induced internalization. *J Biol Chem* **276**, 7968-7973 (2001).
505. Moroder, L. et al. New evidence for a membrane-bound pathway in hormone receptor binding. *Biochemistry* **32**, 13551-13559 (1993).
506. Greasley, P. J., Fanelli, F., Rossier, O., Abuin, L. & Cotecchia, S. Mutagenesis and modelling of the alpha(1b)-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol Pharmacol* **61**, 1025-1032 (2002).
507. Vaidehi, N. et al. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **99**, 12622-12627 (2002).
508. Archer, E., Maigret, B., Escrieut, C., Pradayrol, L. & Fourmy, D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol Sci* **24**,

36-40 (2003).
509. Oliveira, L., Hulsen, T., Lutje Hulsik, D., Paiva, A. C. & Vriend, G. Heavier-than-air flying machines are impossible. *FEBS Lett* **564**, 269-273 (2004).
510. Blanco, F. J., Rivas, G. & Serrano, L. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol* **1**, 584-590 (1994).
511. Blumenstein, M., Matsueda, G. R., Timmons, S. & Hawiger, J. A beta-turn is present in the 392-411 segment of the human fibrinogen gamma-chain. Effects of structural changes in this segment on affinity to antibody 4A5. *Biochemistry* **31**, 10692-10698 (1992).
512. Goudreau, N. et al. NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. *Nat Struct Biol* **1**, 898-907 (1994).
513. Boivin, S. et al. Characterization of Urotensin-II Receptor Structural Domains Involved in the Recognition of U-II, URP, and Urantide. *Biochemistry* **45**, 5993-6002 (2006).
514. Pervaiz, S. & Brew, K. Homology of beta-lactoglobulin, serum retinol-binding protein, and protein HC. *Science* **228**, 335-337 (1985).
515. Bishop, R. E., Penfold, S. S., Frost, L. S., Holtje, J. V. & Weiner, J. H. Stationary phase expression of a novel Escherichia coli outer membrane lipoprotein and its relationship with mammalian apolipoprotein D. Implications for the origin of lipocalins. *J Biol Chem* **270**, 23097-23103 (1995).
516. Flower, D. R., Sansom, C. E., Beck, M. E. & Attwood, T. K. The first prokaryotic lipocalins. *Trends Biochem Sci* **20(12)**, 498-499 (1995).
517. Flower, D. R. The lipocalin protein family: structure and function. *Biochem J* **318**, 1-14 (1996).
518. Skerra, A. Lipocalins as a scaffold. *Biochim Biophys Acta* **1482**, 337-350 (2000).
519. Flower, D. R., North, A. C. & Sansom, C. E. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta* **1482**, 9-24 (2000).
520. Flower, D. R., North, A. C. & Attwood, T. K. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci* **2**, 753-761 (1993).
521. Flower, D. R. Beyond the superfamily: the lipocalin receptors. *Biochim Biophys Acta* **1482**, 327-336 (2000).
522. Shaw, P. H., Held, W. A. & Hastie, N. D. The gene family for major urinary proteins: expression in several secretory tissues of the mouse. *Cell* **32**, 755-761 (1983).
523. Kayser, H. Comprehensive Insect Physiology, Biochemistry and Pharmacology. **10**, 367-415 (1985).
524. Snyder, S. H., Sklar, P. B. & Pevsner, J. Molecular mechanisms of olfaction. *J Biol Chem* **263**, 13971-13974 (1988).
525. Nagata, A. et al. Human brain prostaglandin D synthase has been evolutionarily differentiated from lipophilic-ligand carrier proteins. *Proc Natl Acad Sci U S A* **88**, 4020-4024 (1991).
526. Urade, Y., Nagata, A., Suzuki, Y., Fujii, Y. & Hayaishi, O. Primary structure of rat brain prostaglandin D synthetase deduced from cDNA sequence. *J Biol Chem* **264**, 1041-1045 (1989).
527. Bell, S. C. Secretory endometrial and decidual proteins: studies and clinical significance of a maternally derived group of pregnancy-associated serum proteins. *Hum Reprod* **1**, 129-143 (1986).
528. Bell, S. C. Purification of human secretory pregnancy-associated endometrial alpha 2-globulin (alpha 2-PEG) from cytosol of first trimester pregnancy endometrium. *Hum Reprod* **1**, 313-318 (1986).
529. Flower, D. R. The lipocalin protein family: a role in cell regulation. *FEBS Lett* **354**, 7-11 (1994).
530. Pugsley, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**, 50-108 (1993).
531. Bishop, R. E. The bacterial lipocalins. *Biochim Biophys Acta* **1482**, 73-83 (2000).
532. Campanacci, V. et al. The crystal structure of the Escherichia coli lipocalin Blc suggests a possible role in phospholipid binding. *FEBS Lett* **562**, 183-188 (2004).
533. Peitsch, M. C. & Boguski, M. S. Is apolipoprotein D a mammalian bilin-binding protein? *New Biol* **2**, 197-206 (1990).
534. Yang, C. Y. et al. Structure of human apolipoprotein D: locations of the intermolecular and intramolecular disulfide links. *Biochemistry* **33**, 12451-12455 (1994).
535. Sanchez, D., Ganfornina, M. D. & Bastiani, M. J. Lazarillo, a neuronal lipocalin in grasshoppers with a role in axon guidance. *Biochim Biophys Acta* **1482**, 102-109 (2000).
536. Boyles, J. K., Notterpek, L. M. & Anderson, L. J. Accumulation of apolipoproteins in the regenerating and remyelinating mammalian peripheral nerve. Identification of apolipoprotein D, apolipoprotein A-IV, apolipoprotein E, and apolipoprotein A-I. *J Biol Chem* **265**, 17805-17815 (1990).

537. Nygren, P. A. & Uhlen, M. Scaffolds for engineering novel binding sites in proteins. *Curr Opin Struct Biol* **7**, 463-469 (1997).
538. Padlan, E. A. Anatomy of the antibody molecule. *Mol Immunol* **31**, 169-217 (1994).
539. Lesk, A. M., Branden, C. I. & Chothia, C. Structural principles of alpha/beta barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139-148 (1989).
540. Cowan, S. W., Newcomer, M. E. & Jones, T. A. Crystallographic refinement of human serum retinol binding protein at 2A resolution. *Proteins* **8**, 44-61 (1990).
541. Newcomer, M. E. et al. The three-dimensional structure of retinol-binding protein. *EMBO J* **3**, 1451-1454 (1984).
542. Huber, R. et al. Crystallization, crystal structure analysis and preliminary molecular model of the bilin binding protein from the insect Pieris brassicae. *J Mol Biol* **195**, 423-434 (1987).
543. Huber, R. et al. Molecular structure of the bilin binding protein (BBP) from Pieris brassicae after refinement at 2.0 A resolution. *J Mol Biol* **198**, 499-513 (1987).
544. Bocskei, Z. et al. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **360**, 186-188 (1992).
545. Newcomer, M. E. Structure of the epididymal retinoic acid binding protein at 2.1 A resolution. *Structure* **1**, 7-18 (1993).
546. Brownlow, S. et al. Bovine beta-lactoglobulin at 1.8 A resolution--still an enigmatic lipocalin. *Structure* **5**, 481-495 (1997).
547. Rouvinen, J. et al. Probing the molecular basis of allergy. three-dimensional structure of the bovine lipocalin allergen Bos d 2. *J Biol Chem* **274**, 2337-2343 (1999).
548. Müller, H. N. & Skerra, A. Grafting of a high-affinity Zn(II)-binding site on the beta-barrel of retinol-binding protein results in enhanced folding stability and enables simplified purification. *Biochemistry* **33**, 14126-14135 (1994).
549. Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci U S A* **96**, 1898-1903 (1999).
550. Sidhu, S. S. & Koide, S. Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol* **17**, 481-487 (2007).
551. Schlehuber, S., Beste, G. & Skerra, A. A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J Mol Biol* **297**, 1105-1120 (2000).
552. Schlehuber, S. & Skerra, A. Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov Today* **10**, 23-33 (2005).
553. Gram, H. C. Über die isolierte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. *Fortschritte der Medizin* **2**, 185-189 (1884).
554. Holtje, J. V. Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiol Mol Biol Rev* **62**, 181-203 (1998).
555. Kanemasa, Y., Akamatsu, Y. & Nojima, S. Composition and turnover of the phospholipids in Escherichia coli. *Biochim Biophys Acta* **144**, 382-390 (1967).
556. Kamio, Y. & Nikaido, H. Outer membrane of Salmonella typhimurium: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium. *Biochemistry* **15**, 2561-2570 (1976).
557. Smit, J., Kamio, Y. & Nikaido, H. Outer membrane of Salmonella typhimurium: chemical analysis and freeze-fracture studies with lipopolysaccharide mutants. *J Bacteriol* **124**, 942-958 (1975).
558. Braun, V. Covalent lipoprotein from the outer membrane of Escherichia coli. *Biochim Biophys Acta* **415**, 335-377 (1975).
559. Raetz, C. R. Biochemistry of endotoxins. *Annu Rev Biochem* **59**, 129-170 (1990).
560. Koebnik, R., Locher, K. P. & Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37**, 239-253 (2000).
561. Sankaran, K. & Wu, H. C. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem* **269**, 19701-19706 (1994).
562. Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in E. coli. *Biochim Biophys Acta* **1694**, IN1-9 (2004).
563. Bell, R. M., Mavis, R. D., Osborn, M. J. & Vagelos, P. R. Enzymes of phospholipid metabolism: localization in the cytoplasmic and outer membrane of the cell envelope of Escherichia coli and Salmonella typhimurium. *Biochim Biophys Acta* **249**, 628-635 (1971).
564. White, D. A., Albright, F. R., Lennarz, W. J. & Schnaitman, C. A. Distribution of phospholipid-synthesizing enzymes in the wall and membrane subfractions of the envelope of Escherichia coli. *Biochim Biophys Acta* **249**, 636-642 (1971).

565. Osborn, M. J., Gander, J. E. & Parisi, E. Mechanism of assembly of the outer membrane of Salmonella typhimurium. Site of synthesis of lipopolysaccharide. *J Biol Chem* **247**, 3973-3986 (1972).
566. Jones, N. C. & Osborn, M. J. Translocation of phospholipids between the outer and inner membranes of Salmonella typhimurium. *J Biol Chem* **252**, 7405-7412 (1977).
567. Sperandeo, P. et al. Characterization of lptA and lptB, two essential genes implicated in lipopolysaccharide transport to the outer membrane of Escherichia coli. *J Bacteriol* **189**, 244-253 (2007).
568. Bos, M. P., Robert, V. & Tommassen, J. Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**, 191-214 (2007).
569. Tefsen, B., Geurtsen, J., Beckers, F., Tommassen, J. & de Cock, H. Lipopolysaccharide transport to the bacterial outer membrane in spheroplasts. *J Biol Chem* **280**, 4504-4509 (2005).
570. Danese, P. N. & Silhavy, T. J. Targeting and assembly of periplasmic and outer-membrane proteins in Escherichia coli. *Annu Rev Genet* **32**, 59-94 (1998).
571. Driessen, A. J., Manting, E. H. & van der Does, C. The structural basis of protein targeting and translocation in bacteria. *Nat Struct Biol* **8**, 492-498 (2001).
572. Manting, E. H. & Driessen, A. J. Escherichia coli translocase: the unravelling of a molecular machine. *Mol Microbiol* **37**, 226-238 (2000).
573. Mori, H. & Ito, K. The Sec protein-translocation pathway. *Trends Microbiol* **9**, 494-500 (2001).
574. Papanikou, E., Karamanou, S. & Economou, A. Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol* **5**, 839-851 (2007).
575. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663-669 (2007).
576. White, S. H. & von Heijne, G. Transmembrane helices before, during, and after insertion. *Curr Opin Struct Biol* **15**, 378-386 (2005).
577. Chen, R. & Henning, U. A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. *Mol Microbiol* **19**, 1287-1294 (1996).
578. De Cock, H. et al. Affinity of the periplasmic chaperone Skp of Escherichia coli for phospholipids, lipopolysaccharides and non-native outer membrane proteins. Role of Skp in the biogenesis of outer membrane protein. *Eur J Biochem* **259**, 96-103 (1999).
579. Harms, N. et al. The early interaction of the outer membrane protein phoe with the periplasmic chaperone Skp occurs at the cytoplasmic membrane. *J Biol Chem* **276**, 18804-18811 (2001).
580. Walton, T. A., Sandoval, C. M., Fowler, C. A., Pardi, A. & Sousa, M. C. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc Natl Acad Sci U S A* **106**, 1772-1777 (2009).
581. Lazar, S. W. & Kolter, R. SurA assists the folding of Escherichia coli outer membrane proteins. *J Bacteriol* **178**, 1770-1773 (1996).
582. Rouviere, P. E. & Gross, C. A. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev* **10**, 3170-3182 (1996).
583. Sklar, J. G., Wu, T., Kahne, D. & Silhavy, T. J. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev* **21**, 2473-2484 (2007).
584. Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M. & Tommassen, J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**, 262-265 (2003).
585. Voulhoux, R. & Tommassen, J. Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly. *Res Microbiol* **155**, 129-135 (2004).
586. Montoya, M. & Gouaux, E. Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. *Biochim Biophys Acta* **1609**, 19-27 (2003).
587. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189-1192 (2004).
588. Oomen, C. J. et al. Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23**, 1257-1266 (2004).
589. Tamm, L. K., Hong, H. & Liang, B. Folding and assembly of beta-barrel membrane proteins. *Biochim Biophys Acta* **1666**, 250-263 (2004).
590. Basle, A., Rummel, G., Storici, P., Rosenbusch, J. P. & Schirmer, T. Crystal structure of osmoporin OmpC from E. coli at 2.0 A. *J Mol Biol* **362**, 933-942 (2006).
591. Cowan, S. W. et al. Crystal structures explain functional properties of two E. coli porins. *Nature* **358**, 727-733 (1992).
592. Schirmer, T., Keller, T. A., Wang, Y. F. & Rosenbusch, J. P. Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267**, 512-514 (1995).

593. Forst, D., Welte, W., Wacker, T. & Diederichs, K. Structure of the sucrose-specific porin ScrY from Salmonella typhimurium and its complex with sucrose. *Nat Struct Biol* **5**, 37-46 (1998).
594. van den Berg, B., Black, P. N., Clemons, W. M. J. & Rapoport, T. A. Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304**, 1506-1509 (2004).
595. Buchanan, S. K. et al. Crystal structure of the outer membrane active transporter FepA from Escherichia coli. *Nat Struct Biol* **6**, 56-63 (1999).
596. Ferguson, A. D. et al. Structural basis of gating by the outer membrane transporter FecA. *Science* **295**, 1715-1719 (2002).
597. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K. & Welte, W. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282**, 2215-2220 (1998).
598. Chimento, D. P., Mohanty, A. K., Kadner, R. J. & Wiener, M. C. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nat Struct Biol* **10**, 394-401 (2003).
599. Snijder, H. J. et al. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **401**, 717-721 (1999).
600. Vandeputte-Rutten, L. et al. Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. *EMBO J* **20**, 5033-5039 (2001).
601. Vogt, J. & Schulz, G. E. The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. *Structure* **7**, 1301-1309 (1999).
602. Fernandez, C., Hilty, C., Wider, G., Güntert, P. & Wüthrich, K. NMR structure of the integral membrane protein OmpX. *J Mol Biol* **336**, 1211-1221 (2004).
603. Pautsch, A. & Schulz, G. E. Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**, 1013-1017 (1998).
604. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**, 593-656 (2003).
605. Schirmer, T. & Rosenbusch, J. P. Prokaryotic and eukaryotic porins. *Current Opinion in Structural Biology* **1**, 539-545 (1991).
606. Pages, J. M., James, C. E. & Winterhalter, M. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. *Nat Rev Microbiol* **6**, 893-903 (2008).
607. Schwarz, G., Danelon, C. & Winterhalter, M. On translocation through a membrane channel via an internal binding site: kinetics and voltage dependence. *Biophys J* **84**, 2990-2998 (2003).
608. Danelon, C., Brando, T. & Winterhalter, M. Probing the orientation of reconstituted maltoporin channels at the single-protein level. *J Biol Chem* **278**, 35542-35551 (2003).
609. Dutzler, R., Wang, Y. F., Rizkallah, P., Rosenbusch, J. P. & Schirmer, T. Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* **4**, 127-134 (1996).
610. Meyer, J. E. & Schulz, G. E. Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide-protein interactions. *Protein Sci* **6**, 1084-1091 (1997).
611. Scandella, C. J. & Kornberg, A. A membrane-bound phospholipase A1 purified from Escherichia coli. *Biochemistry* **10**, 4447-4456 (1971).
612. Dekker, N. Outer-membrane phospholipase A: known structure, unknown biological function. *Mol Microbiol* **35**, 711-717 (2000).
613. Kingma, R. L. et al. Unusual catalytic triad of Escherichia coli outer membrane phospholipase A. *Biochemistry* **39**, 10017-10022 (2000).
614. Grodberg, J. & Dunn, J. J. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol* **170**, 1245-1253 (1988).
615. Nakata, N. et al. The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci. *Mol Microbiol* **9**, 459-468 (1993).
616. Kukkonen, M. et al. Lack of O-antigen is essential for plasminogen activation by Yersinia pestis and Salmonella enterica. *Mol Microbiol* **51**, 215-225 (2004).
617. Guina, T., Yi, E. C., Wang, H., Hackett, M. & Miller, S. I. A PhoP-regulated outer membrane protease of Salmonella enterica serovar typhimurium promotes resistance to alpha-helical antimicrobial peptides. *J Bacteriol* **182**, 4077-4086 (2000).
618. Bishop, R. E. et al. Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. *EMBO J* **19**, 5071-5080 (2000).
619. Brozek, K. A., Bulawa, C. E. & Raetz, C. R. Biosynthesis of lipid A precursors in Escherichia coli. A membrane-bound enzyme that transfers a palmitoyl residue from a glycerophospholipid to lipid X. *J Biol Chem* **262**, 5170-5179 (1987).

620. Kawasaki, K., Ernst, R. K. & Miller, S. I. 3-O-deacylation of lipid A by PagL, a PhoP/PhoQ-regulated deacylase of Salmonella typhimurium, modulates signaling through Toll-like receptor 4. *J Biol Chem* **279**, 20044-20048 (2004).
621. Heffernan, E. J. et al. Specificity of the complement resistance and cell association phenotypes encoded by the outer membrane protein genes rck from Salmonella typhimurium and ail from Yersinia enterocolitica. *Infect Immun* **62**, 5183-5186 (1994).
622. Koebnik, R. & Kramer, L. Membrane assembly of circularly permuted variants of the E. coli outer membrane protein OmpA. *J Mol Biol* **250**, 617-626 (1995).
623. Sonntag, I., Schwarz, H., Hirota, Y. & Henning, U. Cell envelope and shape of Escherichia coli: multiple mutants missing the outer membrane lipoprotein and other major outer membrane proteins. *J Bacteriol* **136**, 280-285 (1978).
624. Kleinschmidt, J. H. Membrane protein folding on the example of outer membrane protein A of Escherichia coli. *Cell Mol Life Sci* **60**, 1547-1558 (2003).
625. Qu, J., Behrens-Kneip, S., Holst, O. & Kleinschmidt, J. H. Binding regions of outer membrane protein A in complexes with the periplasmic chaperone Skp. A site-directed fluorescence study. *Biochemistry* **48**, 4926-4936 (2009).
626. Pautsch, A. & Schulz, G. E. High-resolution structure of the OmpA membrane domain. *J Mol Biol* **298**, 273-282 (2000).
627. Killmann, H., Benz, R. & Braun, V. Conversion of the FhuA transport protein into a diffusion channel through the outer membrane of Escherichia coli. *EMBO J* **12**, 3007-3016 (1993).
628. Klebba, P. E., Hofnung, M. & Charbit, A. A model of maltodextrin transport through the sugar-specific porin, LamB, based on deletion analysis. *EMBO J* **13**, 4670-4675 (1994).
629. Charbit, A., Clement, J. M. & Hofnung, M. Further sequence analysis of the phage lambda receptor site. Possible implications for the organization of the lamB protein in Escherichia coli K12. *J Mol Biol* **175**, 395-401 (1984).
630. Killmann, H., Videnov, G., Jung, G., Schwarz, H. & Braun, V. Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. *J Bacteriol* **177**, 694-698 (1995).
631. Morona, R., Klose, M. & Henning, U. Escherichia coli K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. *J Bacteriol* **159**, 570-578 (1984).
632. Beer, K. B. & Miller, V. L. Amino acid substitutions in naturally occurring variants of ail result in altered invasion activity. *J Bacteriol* **174**, 1360-1369 (1992).
633. Kupsch, E. M., Knepper, B., Kuroki, T., Heuer, I. & Meyer, T. F. Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by Neisseria gonorrhoeae for human leukocytes and epithelial cells. *EMBO J* **12**, 641-650 (1993).
634. Prasadarao, N. V. et al. Outer membrane protein A of Escherichia coli contributes to invasion of brain microvascular endothelial cells. *Infect Immun* **64**, 146-153 (1996).
635. Koebnik, R. Membrane assembly of the Escherichia coli outer membrane protein OmpA: exploring sequence constraints on transmembrane beta-strands. *J Mol Biol* **285**, 1801-1810 (1999).
636. Dornmair, K., Kiefer, H. & Jahnig, F. Refolding of an integral membrane protein. OmpA of Escherichia coli. *J Biol Chem* **265**, 18907-18911 (1990).
637. Surrey, T. & Jahnig, F. Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc Natl Acad Sci U S A* **89**, 7457-7461 (1992).
638. Kleinschmidt, J. H., Wiener, M. C. & Tamm, L. K. Outer membrane protein A of E. coli folds into detergent micelles, but not in the presence of monomeric detergent. *Protein Sci* **8**, 2065-2071 (1999).
639. Schweizer, M., Hindennach, I., Garten, W. & Henning, U. Major proteins of the Escherichia coli outer cell envelope membrane. Interaction of protein II with lipopolysaccharide. *Eur J Biochem* **82**, 211-217 (1978).
640. Bonhivers, M. et al. Stability studies of FhuA, a two-domain outer membrane protein from Escherichia coli. *Biochemistry* **40**, 2606-2613 (2001).
641. Kleinschmidt, J. H. & Tamm, L. K. Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism. *Biochemistry* **35**, 12993-13000 (1996).
642. Kleinschmidt, J. H. & Tamm, L. K. Time-resolved distance determination by tryptophan fluorescence quenching: probing intermediates in membrane protein folding. *Biochemistry* **38**, 4996-5005 (1999).
643. Kleinschmidt, J. H., den Blaauwen, T., Driessen, A. J. & Tamm, L. K. Outer membrane protein

A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **38**, 5006-5016 (1999).
644. Popot, J. L. & Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031-4037 (1990).
645. Jansen, C., Heutink, M., Tommassen, J. & de Cock, H. The assembly pathway of outer membrane protein PhoE of Escherichia coli. *Eur J Biochem* **267**, 3792-3800 (2000).

Chapter II: Properties of the N-terminal domains from Y-receptors probed by NMR spectroscopy

*Chao Zou, Sowmini Kumaran, Reto Walser and Oliver Zerbe**

Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH 8057 Zurich, Switzerland

*Corresponding author:

Oliver Zerbe, Institute of Organic Chemistry, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich / Switzerland, phone: +41-44-6354263, fax: +41-44-6356882, email: oliver.zerbe@oci.uzh.ch

Abstract

Binding of neurohormones from the NPY family to their receptors, the so-called Y receptors, that belong to the superfamily 1b of G-protein coupled receptors, might include transient binding to the N-terminal domains of the receptors. Accordingly we have studied structural features of the N-terminal domains from the Y1, Y2, Y4, and Y5 receptor subtypes (N-Y1, N-Y2, N-Y4, N-Y5). We developed efficient strategies for their recombinant expression. N-Y4 and N-Y1 were expressed as insoluble fusions to enforce accumulation into inclusion bodies, whereas N-Y2 and N-Y5 were expressed as soluble fusion proteins. All N-terminal domains are fully flexible in aqueous buffer. In the presence of phospholipid micelles some stretches within the polypeptides adopt helical conformations, but these are too unstable to be characterized in detail. Using chemical shift mapping techniques interactions of NPY, PYY and PP, the three members of the neurohormone family which are the Y receptors' natural ligands, with N-Y1, N-Y2 and N-Y5 revealed chemical shift changes in all cases, with the largest values being encountered for PP interacting with N-Y1 or N-Y5 both in the presence as well as in the absence of phospholipid micelles. The strength of the interactions, however, is generally weak, and the data also point to non-specific contacts. Previously, in case of the interaction of N-Y4 with PP the contacts were shown to be electrostatic in nature. This work indicates that association of the peptides with the N-terminal domains may generally be part of their binding trajectory.

Keywords: GPCR, Y receptors, membrane proteins, structural biology, SPR

1 Introduction

G-protein coupled receptors (GPCRs) present the pharmacologically most important class of receptors and the most important target for pharmaceutical drugs[1]. Recently, significant progress has been made in structural studies of GPCRs. For example, the structures of bovine rhodopsin, the data on the β1 and β2 -adrenergic receptors and on squid rhodopsin[2,3,4,5] have improved our understanding of this biologically important class of proteins.

Generally the structure of GPCRs can be described as an extracellular N-terminal domain (ranging in size from ten to several thousand residues), which is anchored in the plasmamembrane by 7 transmembrane helices (7TM segment). The latter are interconnected by three intra- and three extracellular loops. The 7TM segment is followed by a cytoplasmic C-terminal domain. While the extracellular N-terminal domain of bovine rhodopsin was surprisingly well-structured and revealed the non-anticipated presence of a short anti-parallel β-

sheet, the corresponding segment of the β-adrenergic receptor could not be traced in the electron maps presumably because of its inherent flexibility.

Previously, we have in detail investigated structural properties of a 41 amino acid fragment corresponding to the N-terminal domain of the human Y4 receptor (N-Y4)[6]. This receptor belongs to a class of GPCRs targeted by neurohormones of the neuropeptide Y family[7,8]. The Y receptors are comprised of four subtypes called Y1, Y2, Y4, and Y5 with Y4 showing high affinity and specificity for the pancreatic polypeptide (PP). While unstructured in solution a short α-helical stretch comprising residues 5 to 10 was observed in the presence of phospholipid micelles for N-Y4.

In this work we now report on our recent studies on structural properties of all other human N-terminal domains from the human Y receptors (for sequences see Figure 1).

```
pNPY: YPSKPDNPGE DAPAEDMARY YSALRHYINL ITRQRY-NH2
pPYY: YPAKPEAPGE DASPEELSRY YASLRHYLNL VTRQRY-NH2
bPP : APLEPEYPGD NATPEQMAQY AAELRRYINM LTRPRY-NH2
       *  *  *    *  *    *    ** * *   ** **

hN-Y1:                  MNSTLFSQVENHSVHSNFSEKNAQLLAFENDDCHLPLAMI
hN-Y2:       MGPIGAEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELIDSTKLIEVQ
hN-Y4:                 MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVM
hN-Y5:      MSFYSKQDYNMDLELDEYYNKTLATENNTAATRNSDFPVWDDYKSSVDDLQ
```

Figure 1: *Sequence alignment of the principal members of the NPY family and of the N-terminal domains from the various Y receptor subtypes.*

Synthetic routes for recombinant production of the polypeptides in isotopically labeled form are described and compared to each other. The N-terminal domains from all Y receptors are fully unstructured in aqueous solution. In contrast, in the presence of phospholipid micelles all N termini except of N-Y2 form helical segments with variable degree of stability.

In our previous work we demonstrated that N-Y4 interacts with PP. Surface plasmon resonance (SPR) measurements indicated weak (Kd 50 μM) binding, and subsequent mutagenesis experiments revealed that electrostatic interactions from anionic ligand and cationic N-Y4 residues contributed to that interaction. In this work we also tested binding of the principal members of the NPY family (the neuropeptide Y (NPY), the pancreatic polypeptide (PP) and the peptide YY (PYY)) to all other N-terminal domains from this class of GPCRs.

2 Materials and Methods

$^{15}NH_4Cl$ was from Spectra Isotopes (Columbia, USA),d_{13}-MES, d_{38}-DPC- (99%-d), and D_2O was from Cambridge Isotope Laboratories (Andover, Massachusetts, USA). 5-doxylstearic acid was from Aldrich (Buchs, Switzerland). Oligonucleotide primers were synthesized by Microsynth GmbH (Balgach, Switzerland).

2.1 Expression and purification of N-terminal domains from the human Y receptors

Depending on their stability against proteolysis the N-terminal domains were either expressed as fusions to ubiquitin (N-Y2 and N-Y5) or to ketosteroidisomerase (N-Y1 and N-Y4).

In case of N-Y2 and N-Y5 the amino acid sequence was reverse translated into a DNA sequence taking into account the preferred *E.coli* codon usage including a terminal stop codon and a *Sal*I restriction site. The resulting fragments were purified by electrophoresis and gel extraction and digested with *Sal*I, resulting in fragments that were blunt-ended on one side and contained *Sal*I-cohesive end on the other end. These fragments were ligated into the pUBK19 vector (gift from T. Kohno, Mitsubishi Kasei Institute of Life Science, Tokyo, Japan), which had been digested with *Nsi*I and *Sal*I and purified before. The resulting plasmids were sequenced and transformed into C41 cells[9]. For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done on LB medium. In each case 1 liter of medium containing 50 μg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} around 0.5 with 0.4 mM IPTG. LB- and minimal medium cultures were grown under induction for 4 h and 11 h, respectively. Cells were harvested by centrifugation on a Sorvall GSA rotor at 4 °C and stored at -20 °C. The cell pellets were thawed on ice for 1 h and resuspended in 25 ml denaturing basic buffer (50 mM Tris pH 8, 6 M GdnHCl, 100 mM NaCl, 1 mM β–mercaptoethanol). The suspension was lysed by sonication on ice.

The ubiquitin fusion proteins were purified by Ni-NTA chromatography. Refolding was achieved by applying a linear gradient to exchange the denaturing basic buffer to native binding buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 20 mM imidazole), and the protein was eluted with binding buffer containing 200 mM imidazole. The eluates were diluted 10-fold with basic buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol) and a 1 mg/ml YUH-solution (for expression and purification of YUH see appendix G) was added in a

20-fold dilution. The cleavage reactions were allowed to proceed for 3 hours at 37 °C.

In case of N-Y1 and N-Y4 the DNA sequences were subcloned from wt cDNA of the corresponding Y receptor (University of Missouri-Rolla (UMR) cDNA Resource Center) by PCR. During PCR, a GSGSGS linker followed by TEV cleavage sequence was introduced at the N terminus of the target sequence. After digestion with XhoI and EspI, the fragments were ligated with T4 DNA ligase into the pET31b vector, which had been digested with XhoI and EspI. The correctness of the constructs was verified by DNA sequencing (Synergene Biotech, Switzerland). The resulting plasmids were transformed into BL21(DE3) for expression. For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done in LB medium. In each case 1 liter of medium containing 50 μg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} of 0.7 with 1 mM IPTG, harvested after 5 hours by centrifugation on a Sorvall GSA rotor at 4 °C and the pellets were stored at -20 °C.

The fusion proteins were purified from inclusion bodies by Ni-NTA chromatography in presence of 6 M GdnHCl. After removal of GdnHCl by dialysis the precipitated fusion protein was solubilized in 50 mM Tris pH 8.0 in the presence of 2% N-lauryl sarcosine upon sonication to a final concentration of 2 mg/ml. The resulting solution was dialyzed against a 20-fold excess of 50 mM Tris, pH 8.0 for 4-6 times. The solution was diluted 10 times with 50 mM Tris pH 8.0 and EDTA and DTT were added to a final concentration of 0.5 mM and 1 mM, respectively. TEV protease (for expression and purification of TEV protease see appendix H) was added to a final concentration of 100 mM and the cleavage mixture was incubated at 4°C over night.

All target peptides were finally purified by C18-RP-HPLC (Vydac, USA) by using a water/acetonitrile/0.1% TFA gradient. Yields ranged from 3 mg to 20 mg peptide from 1 liter of culture. The mass of all peptides was confirmed by MALDI-TOF MS or ESI MS: N-Y1: 4532.9 Da (theoretical value: 4533.0 Da); ^{15}N-N-Y1: 4587.0 Da (theoretical value: 4587.0 Da); N-Y2: 5509.3 Da (theoretical value: 5510.0 Da); ^{15}N-N-Y2: 5568.0 Da (theoretical value: 5570.0 Da); N-Y4: 4554.0 Da (theoretical value: 4556.1 Da); ^{15}N-N-Y4: 4614.0 Da (theoretical value: 4611.1 Da); N-Y5: 6053.7 (theoretical value: 6053.4); ^{15}N-N-Y5: 6119.5 Da (theoretical value: 6118.4 Da).

2.2 NMR and CD spectroscopy

For studies of structure or backbone dynamics 1 mM solution of the peptides at pH 5.6, 20 mM d_{13}-MES, 300 mM d_{38}-DPC were used. All spectra were recorded on an AV-700 Bruker NMR

spectrometer at 310 K. Chemical shifts were calibrated to the water line at 4.63 ppm and nitrogen shifts were referenced indirectly to liquid NH_3. The spectra were processed using the Bruker Topspin2.0 software and transferred into the XEASY[10] and Cara[11] programs for further analysis.

For chemical shift assignments 3D ^{15}N-resolved TOCSY and NOESY[12] were used. In case of N-Y5 we decided to use ^{13}C,^{15}N labeling in combination with experiments that directly correlate sequential amide moieties[13]. Upper-distance limits for structure calculations of N-Y1 were derived from a 70 ms NOESY spectrum[14]. Structures were calculated in the program CYANA using its standard simulated annealing protocol[15].

A proton-detected version of the steady-state ^{15}N{^{1}H}-heteronuclear Overhauser effect sequence was used for measurement of the heteronuclear NOE[16]. Therein, the buildup of the NOE was achieved through a pulse train of 120 degree proton pulses separated by 5 ms over a period of 3 seconds.

For measurements of interactions by chemical shift mapping methodology 0.1 mM solutions of the ^{15}N-labeled neurohormones were mixed with the corresponding peptides from the N-terminal domains at pH 5.6, 20 mM d_{13}-MES, 300 mM d_{38}-DPC, and the deviations of peak positions were extracted from the [^{15}N,^{1}H]-HSQC spectra and computed according to $\Delta\delta=SQRT(\Delta(^1H)^2+0.2*\Delta(^{15}N)^2)$. Particular care was taken to ensure that no shifts in pH occurred when adding the N-Y peptides. In case of addition of various equivalents of pNPY to ^{15}N-labeled N-Y2 in the presence of DPC micelles the sample was prepared in 20 mM d_{13}-MES, 300 mM d_{38}-DPC at pH 5.6 and pNPY was added as a solid.

For CD analysis, a certain amount of peptides was dissolved in 300 mM DPC buffered with 20 mM MES (pH 5.6), such that the far UV absorption was around 1. CD spectra were recorded at 37°C on Jasco model J-810 using a quartz cuvette with path length of 1 mm to minimize absorption by the detergent. All spectra were averaged from 3 consecutive measurements in the range between 190 and 250 nm with a slit width of 1 nm and a scanning rate of 5 nm/min. The blank sample was recorded under identical conditions and subtracted from the sample spectra. The final CD intensity is expressed as mean residue ellipticity (deg cm^2 dmol^{-1}).

3 Results

3.1 Expression of N-terminal domains in isotopically labeled form

Isotope labeling of the investigated peptides was required for the study of backbone dynamics using ^{15}N relaxation and for chemical shift mapping experiments for the study of macromolecular interactions. Such labeling precludes the usage of peptides produced from solid phase

synthesis[17,18] and necessitates recombinant production. For reasons of simplicity we generally prefer *E.coli* as the expression host[19]. To avoid rapid degradation in *E.coli*, the peptides need to be linked to a (more) stable fusion partner[20]. Specific cleavage from the fusion partner can be accomplished for systems for which a specific hydrolase (e.g. a ubiquitin hydrolase) is available or by introducing a unique cleavage site, either a protease-sensitive site or a site prone to chemical cleavage such as CNBr[21] or hyroxylamine[22] (see Fig. 2).

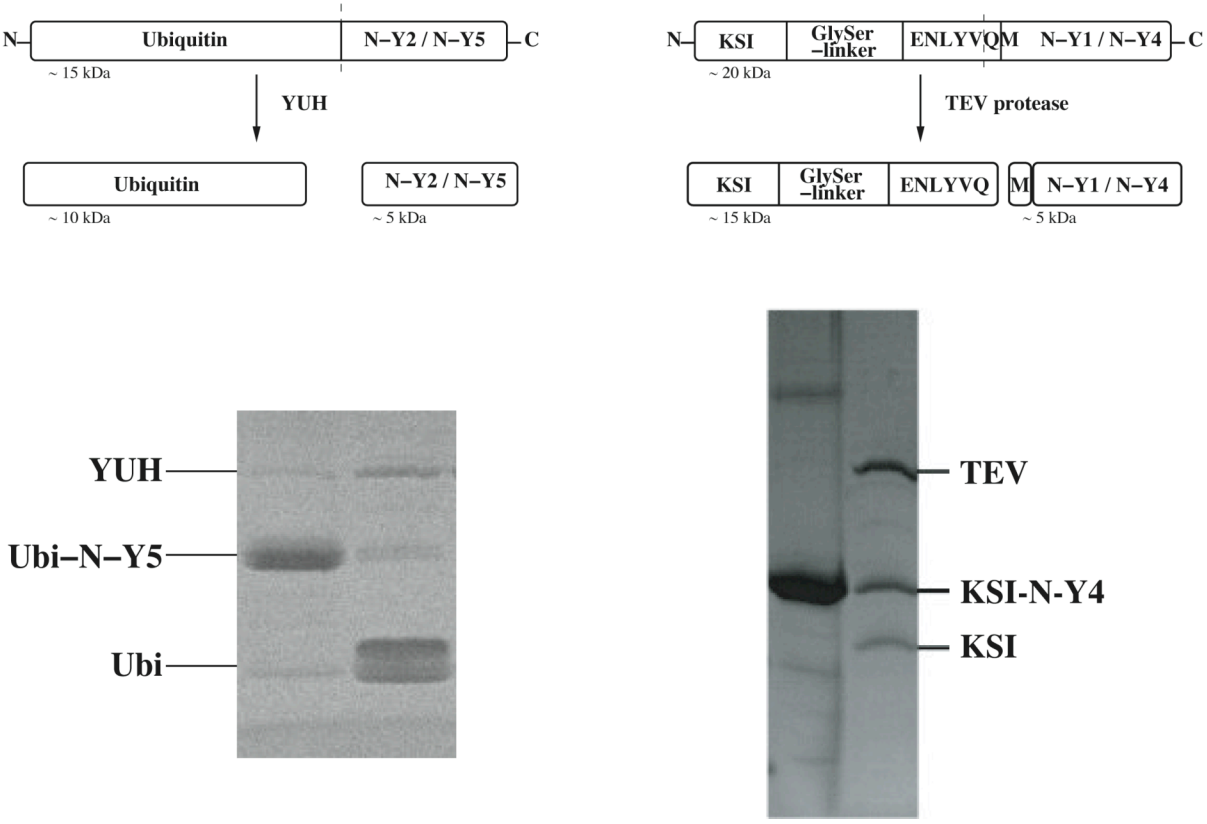


Figure 2: *Scheme showing the two strategies used to produce peptides corresponding to the N-terminal domains of the Y receptors and examples from N-Y5 and N-Y4 for the corresponding scheme, respectively.*

CNBr cleavage in our case was incompatible with the presence of Met residues, and poor efficiency was observed with hydroxylamine, and therefore enzymatic cleavage had to be used. However the latter methods require that the fusion protein can be solubilized under conditions that are compatible with enzymatic activity.

Since the four Y receptor N-terminal fragments studied herein are all reasonably water-soluble and contain Met residues we initially decided to express them in ^{15}N-labeled form as C-terminal fusions to N-terminally decahistidine-tagged yeast ubiquitin[23]. After purification of the fusion construct by Ni-affinity chromatography the desired peptide was liberated through treatment with yeast ubiquitin hydrolase (YUH). This system allowed the recovery of about 6 mg of ^{15}N-labeled N-Y2 and N-Y5 from 1 L of culture. Unfortunately, attempts to express N-Y1 and N-Y4

using this method resulted in unspecific C-terminal degradation. To circumvent intracellular proteolysis, N-Y1 and N-Y4 were expressed as a fusion to the highly water-insoluble protein ketosteroidisomerase (KSI), which resulted in accumulation of the fusion protein in inclusion bodies. A TEV protease cleavage site was introduced between KSI and the target peptide[24,25]. The sequence recognized by the TEV protease is ENLYFQ with Q as the P1' residue. To achieve the natural peptide sequence after cleavage, the P1' residue was replaced with the first residue from the target sequence (here it is Met)[24], and an additional GSGSGS linker was inserted between KSI and the TEV cleavage site to prevent steric hindrance during cleavage.

A problem of the chosen strategy was that the water-insoluble fusion protein must be solubilized in detergent that is compatible with activity of the TEV protease[26]. After extensive detergent screening, we observed that the ionic detergent sarcosyl solubilizes the fusion protein while preserving TEV protease activity to a satisfactory extent. Cleavage efficiency for this system is around 40% allowing recovery of about 2 mg of ^{15}N-labeled N-Y1 and N-Y4 from 1 L of bacterial culture.

3.2 Assignment of chemical shifts

Sequence-specific resonance assignments were done using the strategy developed by Wüthrich and coworkers[27]. Due to extensive resonance overlap of the poorly folded peptides ^{15}N-resolved three-dimensional TOCSY or NOESY data had to be utilized for this task. Representative [^{15}N,^{1}H]-HSQC spectra of all four peptides are depicted in Fig. 3.

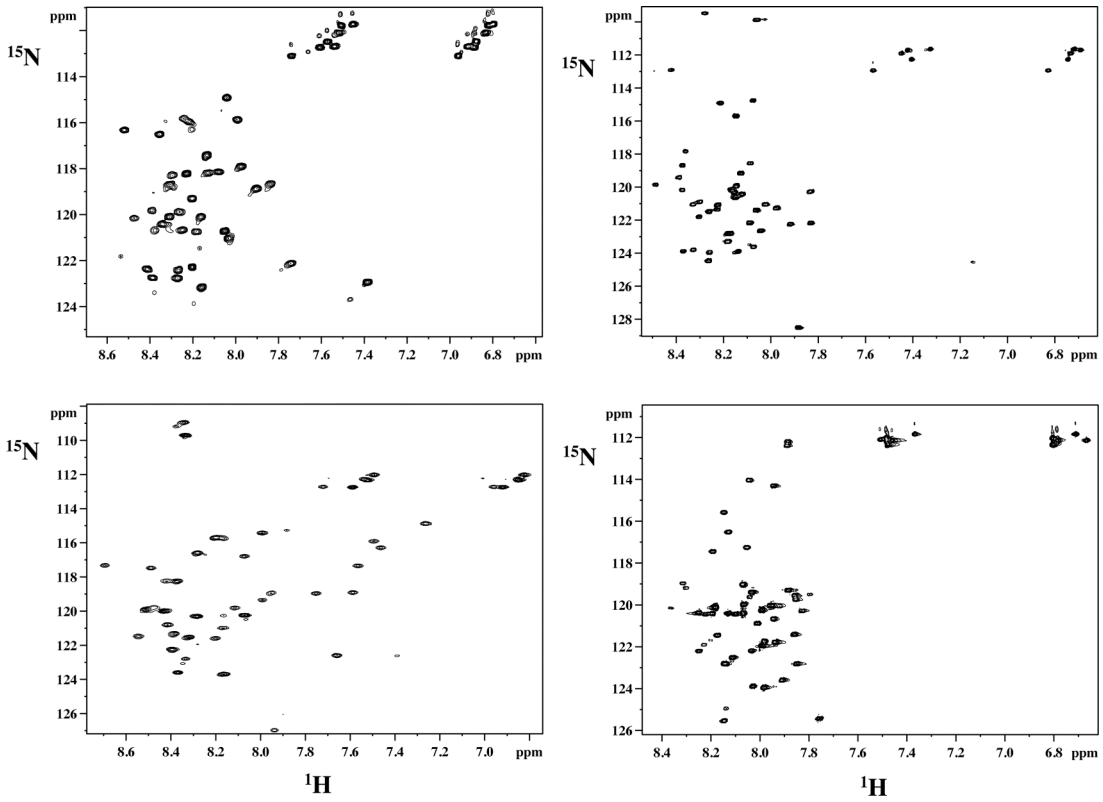


Figure 3: *[^{15}N,^{1}H]-HSQC spectra of all Y receptor N-terminal domains, recorded at 310 K in the presence of DPC micelles. Top left: N-Y1, top right: N-Y2, bottom left: N-Y4, bottom right: N-Y5.*

In case of N-Y5 a ^{13}C,^{15}N-labeled sample, allowing the acquisition of triple resonance spectra, was required. For N-Y1, a set of experiments was first recorded in aqueous buffer. After completed analysis in water the assignments were adjusted to the spectra recorded in the presence of dodecylphosphocholine (DPC) micelles with the help of NOESY spectra. Chemical shifts have been deposited in the BMRB database under accession codes 80.8933262 (N-Y1), 80.6873033 (N-Y2) and 80.74817093 (N-5).

3.3 Screening structural properties using ^{15}N relaxation and CD spectroscopy

CD spectroscopy is a convenient tool to estimate the type and content of secondary structure in peptides and proteins. The CD spectra of all N-terminal domains in the presence of DPC micelles are depicted in Fig 4. The spectrum of N-Y2 displays its minimum around 197 nm, the typical absorption band of unstructured peptides. For all other peptides the minimum is red-shifted and indicates population of helical substructures. The intensities of the absorptions, however, also clearly show that the helical content is very low in all cases, and the typical double minimum at 208 and 222 nm is not visible. For N-Y4, for which we previously observed an α-helix involving residues 5 to 10, the absorption is stronger than for the other peptides.

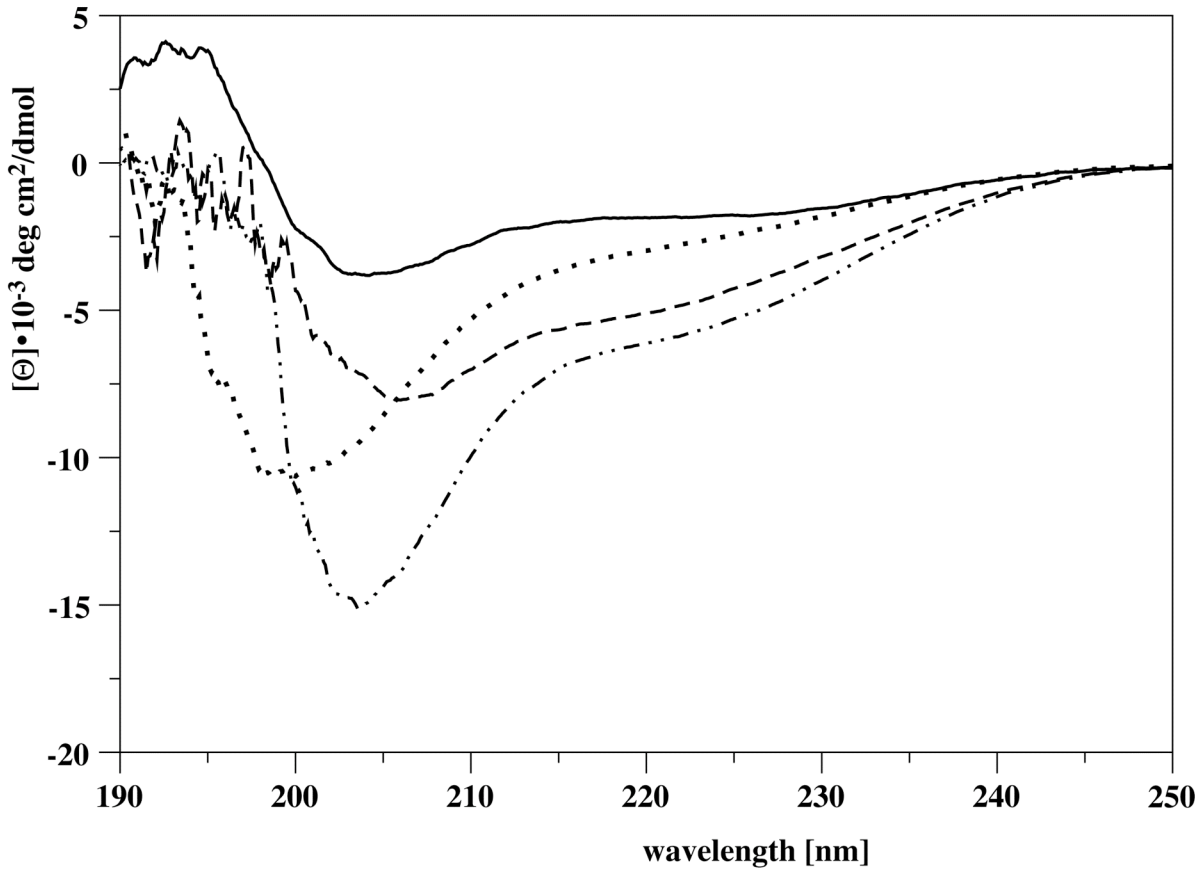


Figure 4: *CD spectra of peptides from all N-terminal domains, recorded at 37 °C in 300 mM DPC, 20 mM MES pH 5.6 solution. Data are shown for N-Y1 (solid line), N-Y2 (dotted line), N-Y4 (dash-dotted line) and N-Y5 (dashed line). Data are converted to mean residue ellipticities.*

The dispersion of the NMR signals in the region of the amide protons is traditionally used to estimate to which extent a peptide or protein is folded[28]. In case of the N-terminal domains from the Y receptors signal dispersion of all peptides was small, indicating that they were largely unfolded. To better assess whether these peptides still contained folded segments we recorded the $^{15}N\{^{1}H\}$-NOEs (H-NOEs). These values range from 0.6 and 0.8 for well-folded elements of secondary structures, and progressively decrease for more flexible amide moieties resulting in negative values for fully flexible segments[29]. The H-NOE data for all N-terminal peptides reveal that all peptides are essentially unstructured in aqueous buffer (data not shown).

Since in the naturally occurring GPCR the N termini are attached to a membrane-protein the backbone dynamics were additionally probed in the presence of a commonly used membrane-mimicking detergent, DPC[30] (see Fig. 5). Again the peptides are not rigidly structured. In the case of N-Y4 we could previously show that a rather stable hydrophobic α-helix is formed between residues 5 and 10, present both in zwitterionic (DPC) as well as in anionic (SDS) micelles[6], reflected by H-NOEs exceeding values of 0.6. In contrast, the N termini from all other Y receptors are less well ordered. The N-Y2 is fully flexible most likely due to the complete lack

of interactions with phospholipid surfaces. The absence of such contacts is supported by the fact that essentially no chemical shift changes occur between N-Y2 in aqueous buffer and in DPC micelles. In contrast, both N-Y1 and N-Y5 reveal short stretches of the polypeptide chain that become rigidified in the presence of the micelles.

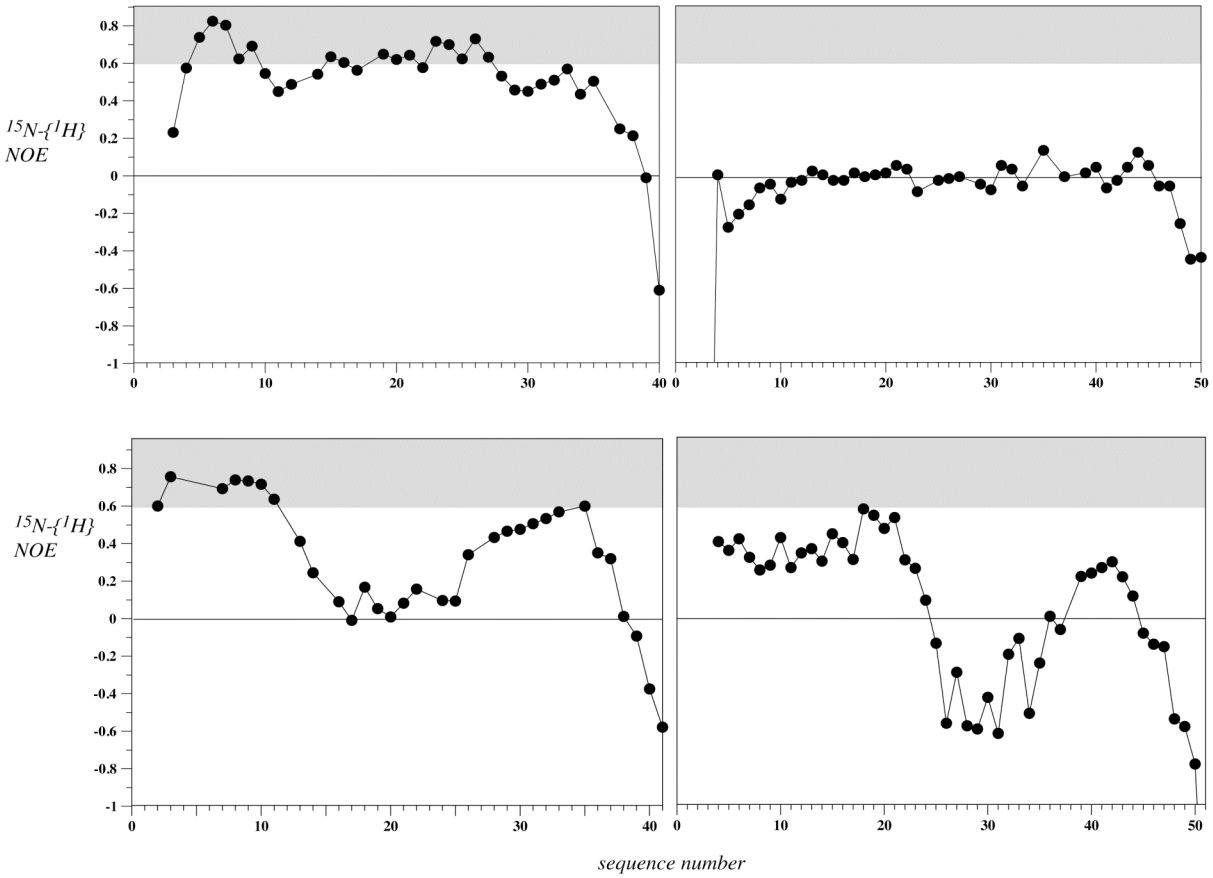


Figure 5: *Values of the ^{15}N-{^{1}H}NOE, recorded at 700 MHz proton frequency along the sequence for Y1 (top left), Y2 (top right), Y4 (bottom left) and Y5 (bottom right). The area containing values larger than 0.6, indicating rather well-folded segments, has been s*

3.4 The structures of the N-terminal domains in the presence of phospholipid micelles

The H-NOE data of N-Y4 revealed the presence of a hydrophobic helix in the segment comprising residues 5 to 10. In addition, a nascent helix was observed in the region including residues 26 to 35. Inspection of the H-NOE data depicted in Fig. 5 clearly indicates that N-Y2 is devoid on any structured segments. Moreover, the H-NOE of N-Y5 is generally below 0.6 and mostly values are even smaller than 0.4. In our experience secondary structure cannot reliably be determined in these cases. We speculate that the molecule, similarly to N-Y4, is segregated into a N-terminal helical region, and a much more destabilized shorter C-terminal helical region separated by a longer non-ordered segment, but the peptide is not ordered sufficiently well to

allow for structural characterization by NMR in detail.

In case of N-Y1, however, elevated values of the H-NOE are observed indicating that this polypeptide may be amenable to more detailed structural studies. Accordingly, we have assigned all proton and nitrogen resonances of N-Y1. During assignment a larger number of contacts involving sequential amide protons were observed, indicating that the ϕ,ψ space of helical backbone conformations was significantly populated. Such stretches were for example observed for residues 4 to 9 and residues 24 to 32. An expansion of the spectral region of the [^{1}H,^{1}H]-NOESY that displays the sequential amide proton NOEs in the segment from 24 to 32 is shown in appendix E. However, except for two α,N (i,i+3) NOEs observed in the segment 4-9 no medium-range contacts were found. The relative strength of intra-residual and sequential αH,NH contacts changes between extended and helical conformations[31], with the intra-residual distance in helices stronger than the sequential one, whereas in extended or unfolded segments the sequential distance is much shorter. A comparison of peak intensities revealed that the sequential NOEs were generally stronger, and in the light of sequential contacts of amide protons, indicate conformational averaging between helical and extended conformations to some extent. Considering this observation it was not really surprising that persistent violations remained in the structure calculations, and helical conformations were only seen involving residues 4 to 9, a region, in which the H-NOE is larger than 0.6. The $^3J(H^N,H^\alpha)$ couplings were larger than 6.5 Hz throughout the sequence (data not shown), reflecting the remaining conformational instability of N-Y1. To our surprise we have not been able to detect any medium-range contacts in the segment 15 to 28, which according to the dynamics data should also be better ordered. We suspect this region to be transiently helical considering the occurrence of sequential amide proton contacts throughout this segment.

To summarize, the spectroscopic data indicate that N-Y4 and N-Y5 are similar in that both contain two helical regions separated by a flexible central segment, with only the N-terminal helix in N-Y4 being well ordered. N-Y1 is largely helical between residues 4 and 28, but the remaining conformational flexibility precludes its detailed structural analysis. N-Y2 is fully flexible and devoid of any detectable residual structure.

3.5 Interaction studies with neuropeptides from the NPY family

We have recently proposed that the peptides of the NPY family may transiently bind to the N-terminal domains of Y receptors in order to become transferred from the membrane-bound state into the genuine binding pocket of the receptor[6,32]. While in that work surface plasmon resonance

was used to establish the strength of the bPP-NY4 (b: bovine) interaction, preliminary experiments using bPP or pPYY (p: porcine) and the N-terminal domains from the other receptors have indicated that the interaction between the peptides and the other N-terminal domains are too weak to be detected by SPR. We have also extensively tried to apply isothermal titration calorimetry but reproducibility of the data recorded in presence of micelles was very unsatisfying, most likely related to the fact that this technique in detergent becomes very challenging when the dissociation constant is more than 10 μM. Therefore we utilized chemical shift mapping experiments both in the presence and absence of DPC micelles in order to derive preliminary data on binding of the peptides from the NPY family to N-Y1, N-Y2 and N-Y5. It should be mentioned, however, that changes in chemical shift To briefly summarize these experiments, we note that in case of pPYY changes are similar (but small) for all N-terminal domains, whereas in case of bPP NY1 or NY5 behave differently compared to N-Y4$^{33\text{-}35}$. To investigate whether pNPY really associates with N-Y2 we have performed a titration experiment, in which up to 10 equivalents of pNPY were added to ^{15}N-labeled N-Y2 (see Fig. 6). The data clearly show concentration-dependent changes of positions of resonances from the N-Y2. Resonances in the segments comprising N-Y2 residues 16-21 and 33-50 are mostly affected. We noticed that acidic residues Glu and Asp are particularly numerous among the residues showing large chemical shift changes when pNPY is added in large excess. This might point to rather non-specific electrostatic contributions to the weak interaction between N-Y2 and pNPY, a fact that has been also observed for the interaction of N-Y4 with bPP. To summarize the interaction studies we can say that significant and reliable effects were only detected in the presence of DPC micelles, and that the interaction of bPP with N-Y4 is much stronger than for the other peptides.

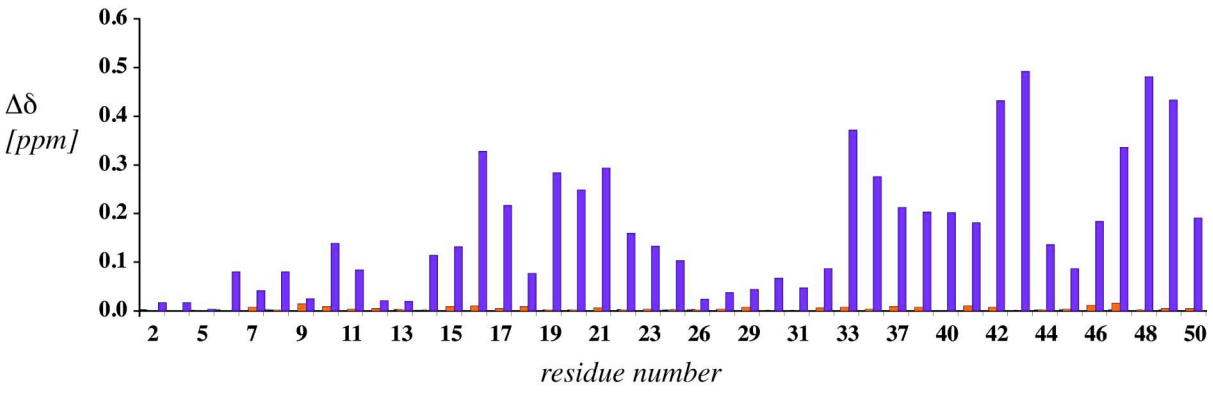


Figure 6: Chemical shift deviation of N-Y2 after addition of 1 and 10 equivalents of pNPY (from left to right). For additional data points at 0.5, 2 and 4 equivalents see appendix I.

4 Discussion

We have postulated, that binding of ligands to Y receptors is preceded by association of the ligands with the plasma membrane. Thereby, the apparent concentration of the ligand is increased and the search for the receptor reduced from three to two dimensions[36-38]. We now studied whether parts of the receptor that protrude into the extracellular compartment may help in transferring ligands, which have accumulated in vicinity of the membrane, into the binding pocket. Such portions of receptors that point into the extracellular space are the N-terminal domains. Herein, we have developed strategies to produce these polypeptides recombinantly in isotopically enriched form for use in high-resolution NMR studies.

The work has demonstrated that these peptides can all be expressed as soluble fusions to ubiquitin. However, N-Y4 and N-Y1 are degraded in the intracellular milieu, and hence much better yields were obtained using insoluble fusions. Cleavage of the target sequence from the insoluble fusion partner could be achieved by solubilizing the fusion protein in the mild detergent sarcosyl, which proved to be compatible with enzymatic activity of the TEV protease used to cleave the peptide from the fusion protein.

Studies on the structure and dynamics of the peptides using NMR revealed that they are all completely disordered in aqueous buffer. In the presence of phospholipid micelles, segments of most receptor N termini became conformationally stabilized, with the exception of N-Y2, which remained unstructured. Otherwise, more (N-Y4) or less stable (N-Y1 or N-Y5) helical segments occurred within the sequences. For all N-terminal peptides chemical shift changes occurred between spectra recorded in presence and absence of DPC micelles, except for N-Y2. This implies that all other peptides associate with the micelle to some extent. Previously, we have made extensive use of the thermodynamic data of Wimley and White for partitioning of single amino acids into the water-membrane interface or the membrane interior[39] to rationalize how peptides interact with phospholipids micelles. A common observation was that the occurrence of the aromatic residues Trp and Tyr help in anchoring peptides in the interface[40]. The partitioning values of the four sequences of the N-terminal domains from the Y receptor subtypes are shown in appendix F. In N-Y4 a stretch comprising residues 5 to 11 is predicted to show partitioning into the micelle interior. This corresponds exactly to the region that becomes helically structured in the presence of micelles. In case of N-Y2 many negatively charges residues occur throughout the sequence, while they are clustered in the central (unstructured) segment in N-Y4. Even more importantly, many Pro residues are present in N-Y2 that might prevent formation of secondary

structure. The sequence of N-Y5 in comparison to N-Y2 is much more amphiphilic in nature, and therefore more likely to favorably interact with the micelles. Again, the regions that become better structured in the presence of DPC micelles correspond to stretches rich in hydrophobic/aromatic residues and hence are predicted to partition into the micelles. The fact that the N-termini of the Y receptors are largely unstructured is compatible with present structural knowledge derived from crystal structures of family 1 GPCRs[2-5,41,42]. In these the N-termini are rather flexible. However, a short two-stranded antiparallel β-sheet complementing the β-sheet formed by residues of the long e2 loop is encountered in bovine rhodopsin[2], whereas a very short helix is present in squid rhodopsin[5]. In both β-adrenergic receptors, the N-terminal portions are not defined in the crystal structures[4,42].

Our interaction studies using chemical shift mapping indicated that bPP strongly interacts with all N-terminal domains, but differences in the sensitivity of certain positions are observed. In contrast, for pPYY or NPY the changes are smaller, and tertiary structure is lost when pPYY associates with the membrane surface. Transient binding of PYY to any of the N-terminal domains is expected to alter the equilibrium between membrane-associated peptide, which is devoid of tertiary structure, and the membrane-detached peptide, that could possibly re-adopt its PP fold. In our BiaCore measurements we could detect strongest binding (Kd approx. 50 μM) for the bPP-NY4 interaction, and chemical shift mapping also revealed the largest changes for bPP upon addition of N-Y4. The fact that the interaction between bPP and N-Y4 is much stronger (Kd 50 μM) than for any other combination of neurohormone and N-terminal domain could indicate that this contact additionally contributes to binding in the pocket, and may help to explain why binding of PP to the Y4 receptor is much tighter than binding of PP to the other subtypes, or stronger than binding of NPY or PYY to all Y receptors. Otherwise we generally see a rather weak non-specific electrostatic interaction. Whether these interactions are strong enough to really help promoting the membrane-bound peptides into the receptor-binding pocket is unclear based on these data. It however is unlikely, that they really significantly contribute to binding in the genuine binding pocket. Very recently, Beck-Sickinger et al. investigated mutants of the Y receptor, in which the N-termini were truncated[43]. Their data indicate little loss of binding affinity and signal transduction in the truncated mutants except for the Y2 receptor. Whether these observations indicate that diffusion of the ligand from the membrane-bond state into the receptor-binding pocket may proceed via different, alternative pathways will need to be subject to further studies.

To summarize this work has described synthetic methods to produce all N-terminal domains in

isotopically labeled form in quantities sufficient for the analysis by various biophysical methods. Structural studies revealed them to be fairly flexible. However, although N-Y2 is fully unfolded, residual helical structures were detected in N-Y1 and N-Y5. For the case of N-Y4 we could previously detect a short rather rigid α-helical stretch in the presence of DPC micelles. In contrast to N-Y4, the nascent helical regions of N-Y1 and N-Y5 contain too much residual motion, so that structure calculations did not fully converge towards α-helical structures. All peptides interact with the N-terminal domains of N-Y1, N-Y2 and N-Y5, but the interactions are weaker than those previously described for bPP binding to N-Y4.

Acknowledgments

We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173) and the Forschungskredit of the University of Zurich (to R.W. and S.K.).

5 References

1. Ma, P. & Zemmel, R. Value of novelty? *Nat Rev Drug Discov* **1**, 571-572 (2002).
2. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
3. Cherezov, V. et al. High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
4. Rasmussen, S. G. et al. Crystal structure of the human β2-adrenergic G-protein-coupled receptor. *Nature* **450**, 383-387 (2007).
5. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363-367 (2008).
6. Zou, C., Kumaran, S., Markovic, S., Walser, R. & Zerbe, O. Studies of the structure of the N-terminal domain from the Y4 receptor - a G protein-coupled receptor - and its interaction with hormones from the NPY family. *Chembiochem* **9**, 2276-2284 (2008).
7. Larhammar, D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **65**, 165-174 (1996).
8. Larhammar, D. & Salaneck, E. Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-151 (2004).

9. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-298 (1996).

10. Ch. Bartels, T.-H. X., M. Billeter, P. Güntert and K. Wüthrich. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **5**, 1-10 (1995).

11. Keller, R. The Computer Aided Resonance Assignment. (2004).

12. Marion, D. et al. Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry* **28**, 6150-6156 (1989).

13. Weisemann, R., Ruterjans, H. & Bermel, W. 3D triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins. *J Biomol NMR* **3**, 113-120 (1993).

14. Kumar, A., Ernst, R. R. & Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* **95**, 1-6 (1980).

15. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378 (2004).

16. Noggle, J. H. & Schirmer, R. E. The Nuclear Overhauser Effect - Chemical Applications. (1971).

17. Fileds, G. B. & Colowick, S. P. Solid-Phase Peptide Synthesis. (1997).

18. Merrifield, B. Concept and early development of solid-phase peptide synthesis. *Methods Enzymol* **289**, 3-13 (1997).

19. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307**, 249-264 (2008).

20. LaVallie, E. R., McCoy, J. M., Smith, D. B. & Riggs, P. Enzymatic and chemical cleavage of fusion proteins. *Curr Protoc Mol Biol* **Chapter 16**, Unit16.4B (2001).

21. Kuliopulos, A. & Walsh, C. Production, Purification, and Cleavage of Tandem Repeats of Recombinant Peptides. *J. Am. Chem. Soc,* **116**, 4599-4607 (1994).

22. Bornstein, P. & Balian, G. Cleavage at Asn-Gly bonds with hydroxylamine. *Methods Enzymol* **47**, 132-145 (1977).

23. Kohno, T., Kusunoki, H., Sato, K. & Wakamatsu, K. A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and ^{15}N/^{13}C. *J Biomol NMR* **12**, 109-121 (1998).

24. Kapust, R. B., Tozser, J., Copeland, T. D. & Waugh, D. S. The P1' specificity of tobacco

etch virus protease. *Biochem Biophys Res Commun* **294**, 949-955 (2002).

25. Kapust, R. B. et al. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993-1000 (2001).

26. Mohanty, A. K., Simmons, C. R. & Wiener, M. C. Inhibition of tobacco etch virus protease activity by detergents. *Protein Expr Purif* **27**, 109-114 (2003).

27. Wüthrich, K. NMR of Proteins and Nucleic Acids. (1986).

28. Wishart, D. S., Sykes, B. D. & Richards, F. M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* **222**, 311-333 (1991).

29. Palmer, A. G. III. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* **30**, 129-155 (2001).

30. Brown, L. R., Bosch, C. & Wüthrich, K. Location and orientation relative to the micelle surface for glucagon in mixed micelles with dodecylphosphocholine: EPR and NMR studies. *Biochim Biophys Acta* **642**, 296-312 (1981).

31. Wüthrich, K., Billeter, M. & Braun, W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J Mol Biol* **180**, 715-740 (1984).

32. Zerbe, O. et al. Recognition of neurohormones of the NPY family by their receptors. *J Recept Signal Transduct Res* **26**, 487-504 (2006).

33. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. Structure and dynamics of micelle-bound neuropeptide Y: comparison with unligated NPY and implications for receptor selection. *J Mol Biol* **305**, 307-329 (2001).

34. Bettio, A., Dinger, M. C. & Beck-Sickinger, A. G. The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci* **11**, 1834-1844 (2002).

35. Cowley, D. J., Hoflack, J. M., Pelton, J. T. & Saudek, V. Structure of neuropeptide Y dimer in solution. *Eur J Biochem* **205**, 1099-1106 (1992).

36. Moroder, L. et al. New evidence for a membrane-bound pathway in hormone receptor binding. *Biochemistry* **32**, 13551-13559 (1993).

37. Sargent, D. F. & Schwyzer, R. Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc Natl Acad Sci U S A* **83**, 5774-5778 (1986).

38. Schwyzer, R. Membrane-assisted molecular mechanism of neurokinin receptor subtype selection. *EMBO J* **6**, 2255-2259 (1987).

39. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3**, 842-848 (1996).

40. Ridder, A. N. et al. Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins. *Biochemistry* **39**, 6521-6528 (2000).

41. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183-187 (2008).

42. Warne, T. et al. Structure of a β1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491 (2008).

43. Lindner, D., Walther, C., Tennemann, A. & Beck-Sickinger, A. G. Functional role of the extracellular N-terminal domain of neuropeptide Y subfamily receptors in membrane integration and agonist-stimulated internalization. *Cell Signal* **21**, 61-68 (2009).

6 Appendix

In tables A-D chemical shifts were referenced to the water line taken at 4.63 ppm at 310K. The ^{15}N scale was derived indirectly by multiplying the frequency of 0 ppm for protons (the Bruker parameter SF) by 0.101329118. Chemical shifts have been deposited in the BMRB data base under deposition codes 80.8933262 (N-Y1), 80.6873033 (N-Y2) and 80.74817093 (N-5).

A - Chemical Shifts of N-Y1 in the presence of DPC micelles

	H^N	H^{α}	H^{β}	others
Met 1	-	-	-, -	γCH_2 -, -; εCH_3 -
Asn 2	-	-	-, -	δNH_2 -, -
Ser 3	8.52	4.42	-, -	γOH -
Thr 4	8.32	4.28	4.01	γCH_3 1.17; γOH -
Leu 5	8.23	4.03	1.86, 1.86	γH 1.17; δCH_3 0.79, 0.72
Phe 6	7.97	4.46	2.96, 3.18	δH 7.19; εH -, - ; ζH -
Ser 7	8.02	4.31	3.86, 3.86	γOH -
Gln 8	8.25	4.28	2.04, 2.04	γCH_2 2.33; εNH_2 7.44
Val 9	7.90	3.94	2.05	γCH_3 0.88
Glu 10	8.26	4.09	1.85, 1.88	γCH_2 2.33; εH -
Asn 11	8.29	4.51	2.69, 2.69	δNH_2 7.58, 6.86
His 12	8.32	4.57	3.05, 3.19	δ^1NH -; δ^2H 7.09; ε^1H -; ε^2NH -
Ser 13	8.18	4.40	3.79, 3.79	γOH -
Val 14	8.15	4.00	2.01	γCH_3 0.81
His 15	8.30	4.61	3.09, 3.09	δ^1NH -; δ^2H 7.05; ε^1H -; ε^2NH -
Ser 16	8.22	4.62	3.77, 3.80	γOH -
Asn 17	8.17	4.19	2.63, 2.63	δNH_2 7.50, 6.81
Phe 18	8.22	4.50	2.99, 3.10	δH 7.17; εH 7.25; ζH -
Ser 19	8.17	4.29	3.79, 3.79	γOH -
Glu 20	8.38	4.14	1.96, 1.96	γCH_2 2.22; εH -
Lys 21	8.15	4.18	1.71, 1.71	γCH_2 1.37; δCH_2 1.94, 2.02; εCH_2 2.74; ζNH_3^+ -
Asn 22	8.13	4.57	2.68, 2.68	δNH_2 7.49, 6.81
Ala 23	8.13	4.11	1.36	
Gln 24	8.13	4.15	2.02, 2.02	γCH_2 2.31; εNH_2 7.49
Leu 25	8.01	4.16	1.64, 1.64	γH 1.54; δCH_3 0.88, 0.81
Leu 26	7.83	4.17	1.57, 1.57	γH 1.46; δCH_3 0.83, 0.79
Ala 27	7.74	4.19	1.23	
Phe 28	7.97	4.55	3.13, 3.13	δH 7.20; εH 7.04 ; ζH -
Glu 29	8.33	4.20	1.87, 1.87	γCH_2 2.18; εH -
Asn 30	8.30	4.62	2.78, 2.78	δNH_2 7.58, 6.86

Asp 31	8.19	4.50	2.58, 2.58	δH -
Asp 32	8.07	4.29	2.78, 2.78	δH -
Cys 33	8.29	4.75	2.59, 2.65	γSH -
His 34	8.36	4.66	3.09, 3.09	δ^1NH -; δ^2H 7.13; ε^1H -; ε^2NH -
Leu 35	8.32	4.51	1.62, 1.62	γH 1.44; δCH_3 0.86
Pro 36		4.44	1.91, 1.96	γCH_2 2.19,2.22 ; δCH_2 3.78, 3.83
Leu 37	8.05	4.18	1.57, 1.57	γH -; δCH_3 0.87, 0.81
Ala 38	8.18	4.27	1.32	
Met 39	8.22	4.38	2.06	γCH_2 2.55, 2.66; εCH_3 1.96
Ile 40	7.38	4.01	1.79	γCH_2 1.09, 1.37; δCH_3 0.82

B - Amide proton and ^{15}N chemical shifts of N-Y1

	N	H^N			N	H^N
Met 1				Lys 21	120.03	8.16
Asn 2				Asn 22	118.11	8.13
Ser 3	116.26	8.52		Ala 23	123.09	8.16
Thr 4	116.44	8.36		Gln 24	117.35	8.13
Leu 5	122.70	8.27		Leu 25	120.96	8.03
Phe 6	115.80	7.99		Leu 26	118.60	7.84
Ser 7	114.84	8.04		Ala 27	122.05	7.74
Gln 8	120.60	8.25		Phe 28	117.83	7.98
Val 9	118.80	7.90		Glu 29	120.35	8.34
Glu 10	122.35	8.27		Asn 30	118.61	8.31
Asn 11	118.23	8.29		Asp 31	120.03	8.31
His 12	119.81	8.26		Asp 32	118.07	8.08
Ser 13	115.76	8.24		Cys 33	120.08	8.47
Val 14	120.68	8.18		His 34	119.74	8.39
His 15	120.62	8.38		Leu 35	122.68	8.39
Ser 16	116.22	8.20		Pro 36		
Asn 17	119.23	8.20		Leu 37	120.63	8.05
Phe 18				Ala 38	122.21	8.20
Ser 19	115.90	8.21		Met 39	118.15	8.23
Glu 20	122.30	8.41		Ile 40	122.86	7.38

C – Amide proton and ^{15}N chemical shifts of N-Y2

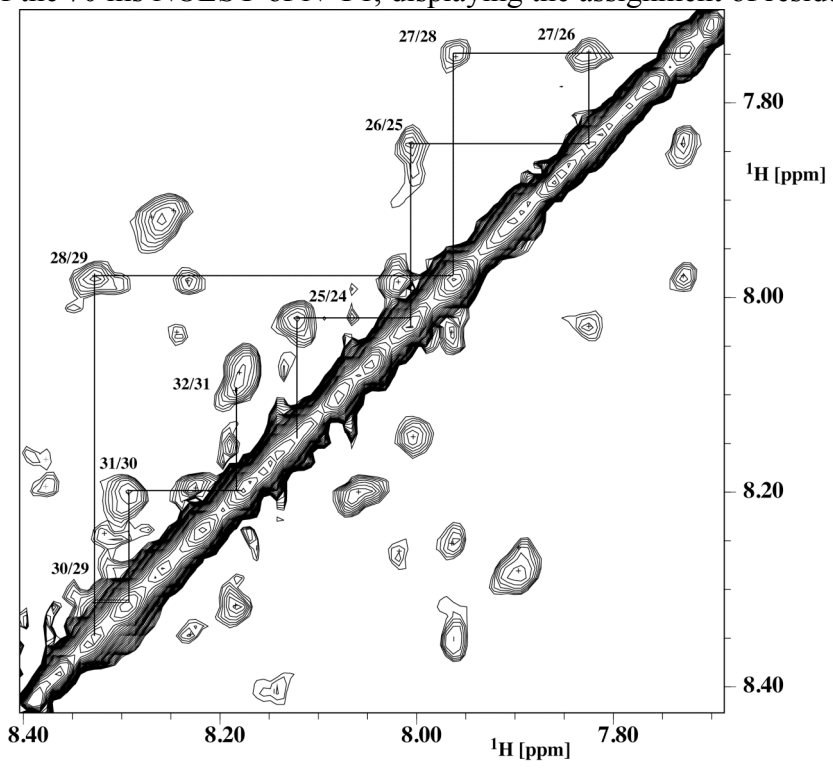
	N	H^N			N	H^N
Met 1				Thr 26	114.7	8.08
Gly 2	112.1	8.59		Thr 27	118.5	8.09
Pro 3				Pro 28		
Ile 4	120.1	8.16		Arg 29	121.0	8.33
Gly 5	112.9	8.42		Gly 30	109.4	8.28
Ala 6	123.6	8.07		Glu 31	119.9	8.15
Glu 7	119.4	8.39		Leu 32	122.8	8.19
Ala 8	123.9	8.14		Val 33	122.6	8.04
Asp 9	119.1	8.12		Pro 34		
Glu 10	120.9	8.30		Asp 35	121.3	8.23
Asn 11	118.7	8.37		Pro 36		
Gln 12	120.3	8.15		Glu 37	121.8	8.30
Thr 13	115.7	8.15		Pro 38		
Val 14	122.2	8.09		Glu 39	120.1	8.37
Glu 15	123.8	8.33		Leu 40	123.3	8.18
Glu 16	121.5	8.26		Ile 41	121.2	7.98
Met 17	121.1	8.22		Asp 42	123.9	8.26
Lys 18	122.8	8.18		Ser 43	117.8	8.36
Val 19	121.0	8.02		Thr 44	114.9	8.21
Glu 20	123.9	8.37		Lys 45	122.2	7.83
Gln 21	120.6	8.15		Leu 46	122.2	7.92
Tyr 22	120.4	8.12		Ile 47	120.2	7.83
Gly 23	109.8	8.06		Glu 48	124.4	8.26
Pro 24				Val 49	121.4	8.06
Gln 25	119.8	8.49		Gln 50	128.5	7.89

D – Amide proton and ^{15}N chemical shifts of N-Y5

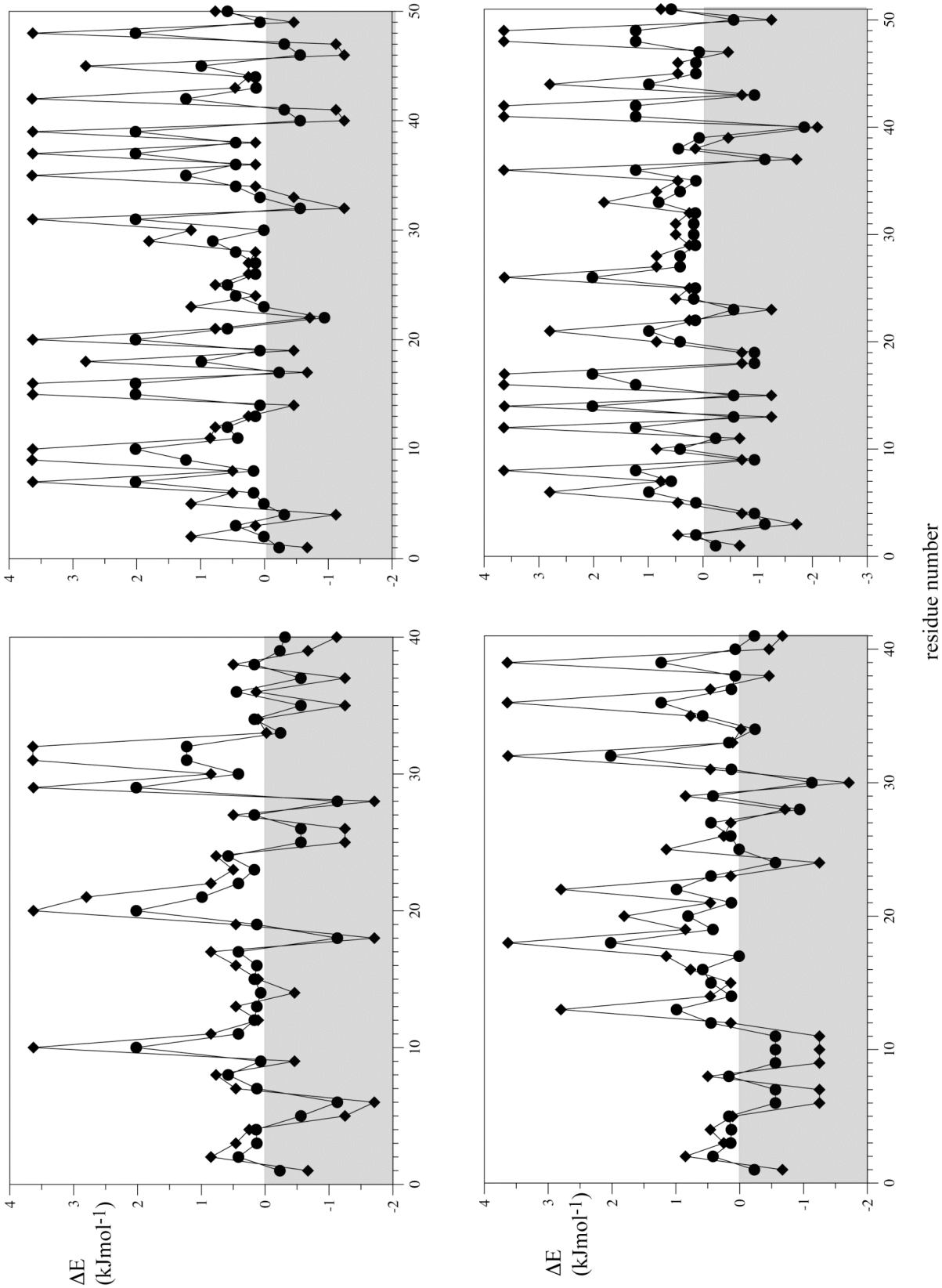
	N	H^N			N	H^N
Met 1				Asn 27	120.23	8.43
Ser 2				Asn 28	119.29	8.35
Phe 3	118.81	7.84		Thr 29	114.05	8.05
Tyr 4	118.03	7.63		Ala 30	125.61	8.17
Ser 5	115.85	8.10		Ala 31	122.26	8.07
Lys 6	122.27	8.25		Thr 32	112.43	7.92
Gln 7	119.04	8.18		Arg 33	122.34	8.32
Asp 8	119.37	8.03		Asn 34	119.25	8.36
Tyr 9	120.84	8.43		Ser 35	115.51	8.17
Asn 10	119.60	8.25		Asp 36	121.69	8.26
Met 11	120.23	8.22		Phe 37	120.00	8.11
Asp 12	120.96	8.38		Pro 38		
Leu 13	121.09	7.90		Val 39	117.15	8.04
Glu 14	120.56	8.39		Trp 40	122.72	8.12

Leu 15	121.51	8.39		Asp 41	118.58	8.12
Asp 16	118.01	8.47		Asp 42	118.49	7.95
Glu 17	118.35	8.04		Tyr 43	119.01	7.82
Tyr 18	117.72	7.86		Lys 44	121.02	7.75
Tyr 19	118.50	8.10		Ser 45	115.33	8.12
Asn 20	117.52	8.21		Ser 46	117.18	8.06
Lys 21	118.77	7.90		Val 47	118.82	7.84
Thr 22	113.27	7.80		Asp 48	122.02	8.04
Leu 23	121.69	7.83		Asp 49	119.12	8.03
Ala 24	122.63	7.76		Leu 50	121.18	7.90
Thr 25	112.49	7.92		Gln 51	124.45	7.67
Glu 26	122.74	8.19				

E - Expansion of the 70 ms NOESY of N-Y1, displaying the assignment of residues 24 to 31:



F - Free energies for partitioning residues of the Y receptor N-terminals (top left: N-Y1, top right: N-Y2, bottom left: N-Y4 and bottom right: N-Y5) domains into the water-membrane interface (circles) or into the membrane interior (diamonds) (data taken from Wimley, Nat Struct Biol, **3**, 842-848, 1996). Regions of favorable values are shaded in gray.



G - Expression and purification of YUH:

Yeast ubiquitin hydrolase (YUH) was expressed with a C-terminal hexahistidine tag and purified on a Ni-NTA column. The plasmid coding for the YUH-construct pYUHK20b was a generous gift from Toshiyuki Kohno (Mitsubishi Kasei Institute of Life Science, Tokyo, Japan).

5 ml of LB-broth containing 50 μg/ml kanamycin were inoculated with a colony of BL21 DE3 pYUHK20b cells, streaked onto plate from a glycerol stock, and incubated for 12 h at 37 ˚C and 220 rpm. 0.5 liter of LB-broth containing 50 μg/ml kanamycin were inoculated with the 5 ml overnight culture and incubated at 37 ˚C and 240 rpm in a 2 l Erlenmeyer flask. The culture was induced with 0.4 mM IPTG at an OD_{600} of 0.6-0.7 and grown for another 5 h to a final OD_{600} of around 5. The culture was harvested by centrifugation at 5000 rpm and 4 ˚C for 20 min. 7 g of wet biomass were obtained from 1 l of culture. The cell pellet was frozen at -20 ˚C.

The cell pellet was thawed on ice and resuspended in 40 ml resuspension buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol). 9 mg of lysozyme were added and the mixture incubated on ice for 15 min. The resuspension mixture was sonicated on ice with a Branson Digital Sonifier.

The lysate was centrifuged twice at 190000 rpm and 4 ˚C for 45 min and loaded onto a 10 ml column volume (CV) Ni-NTA-agarose column previously equilibrated with running buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 10 mM imidazole). Bound protein was eluted with elution buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 100 mM imidazole). The eluate was confirmed to contain the target protein by SDS-PAGE. To 10 ml of eluate 1.1 ml of glycerol were added to yield a final glycerol concentration of 10%. The YUH-concentration of this mixture was determined by a Bradford assay to be 7 mg/ml. This solution was stored at -20 °C in 1 ml aliquots.

H - Expression and purification of TEV protease:

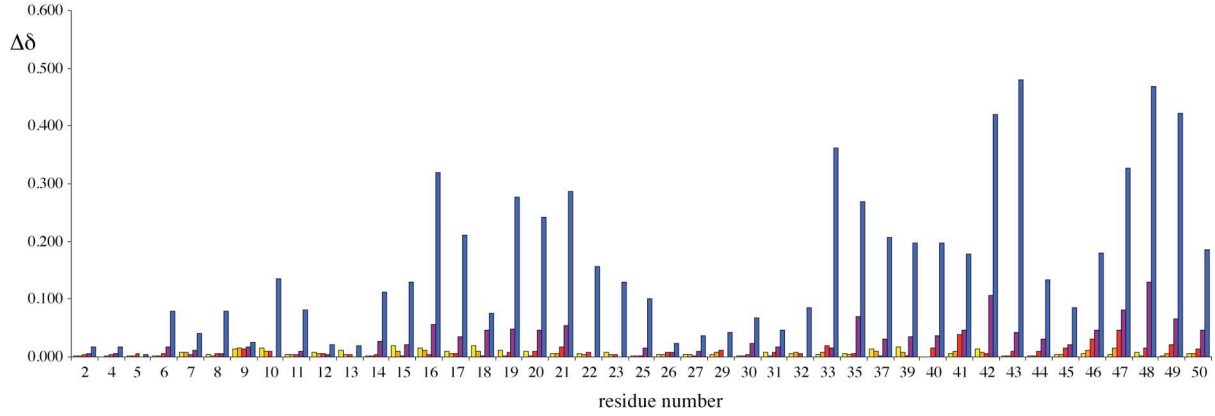
The plasmid pTH24[1] was transformed to Rosetta(DE3) pLys cells. 0.5 ml of an overnight LB culture containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol were used to inoculate 1 liter of TB also containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The culture was incubated at 37 °C. When it reached an OD_{600} of 0.6 it was induced with 0.1 mM IPTG and the temperature was lowered to 20 °C. After 20 hours the cells were harvested by centrifugation and the cell pellet was stored at –20 °C.

The cell pellet from 1 liter culture was resuspended in 40 ml washing buffer (50 mM

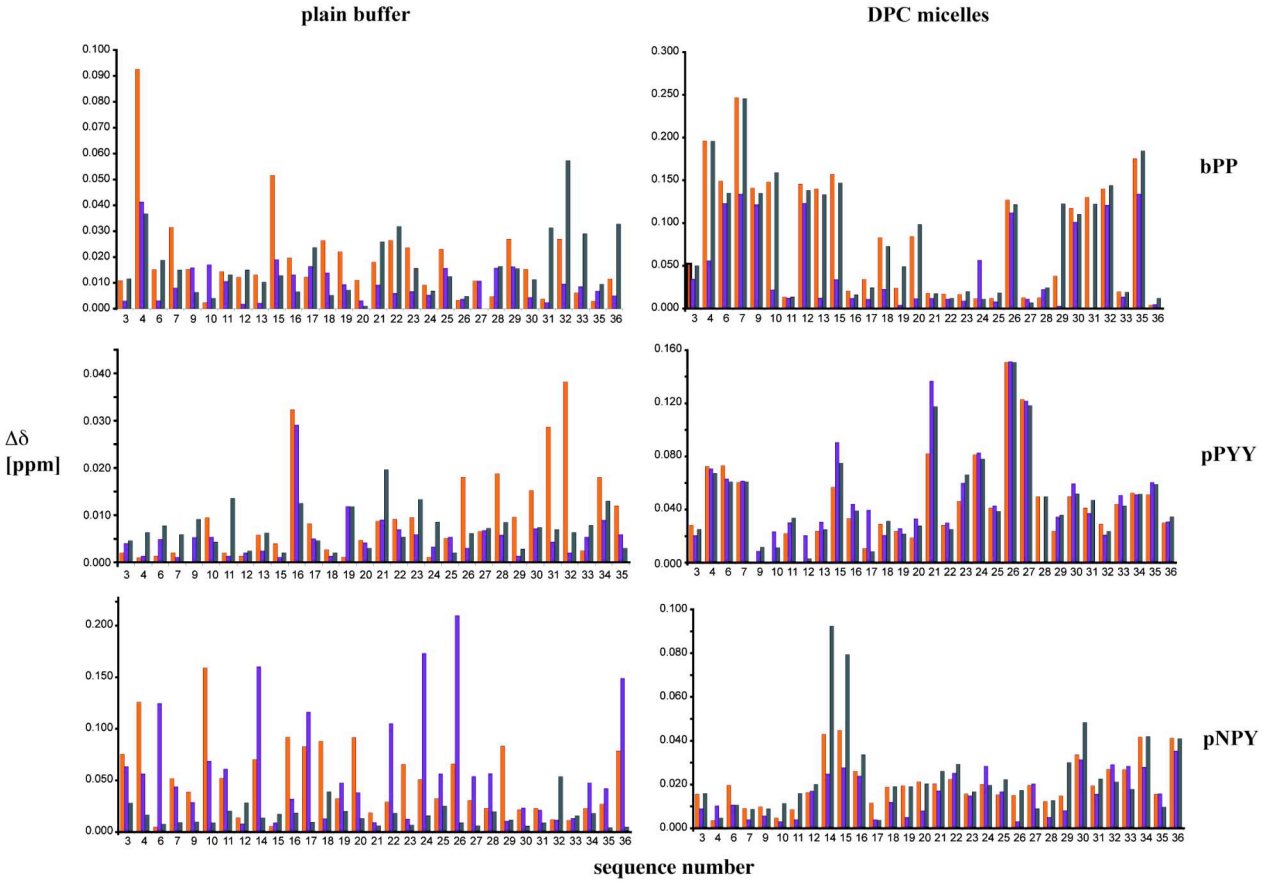
sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7). The resuspension mixture was sonicated on ice with a Branson Digital Sonifier. After centrifugation at 4 °C at 30000 g for 30 min the supernatant was loaded onto a Ni-NTA column. Unbound protein was eluted with washing buffer and bound protein was eluted with elution buffer (washing buffer with 200 mM imidazole). EDTA and DTT were added to a final concentration of 2 and 10 mM, respectively. 10 ml eluate were dialyzed over night against 1 liter dialysis buffer (25 mM sodium phosphate, 200 mM NaCl, 2 mM EDTA, 2 mM DTT, pH 8) at 4 °C. 10% glycerol were added and the protease solution was stored at –20 °C.

1) Susanne van den Berg, Per-Ake Lofdahl, Torleif Hard, Helena Berglund, Improved solubility of TEV protease by directed evolution, 2006, J. Biotechnol., 121, 291-298.

I - Titration of N-Y2 with 0.5 (yellow), 1 (orange), 2 (red), 4 (purple), and 10 (blue) equiv. of pNPY.

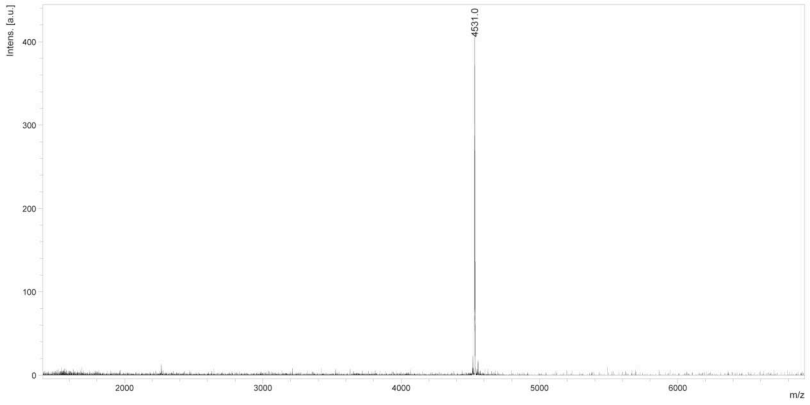


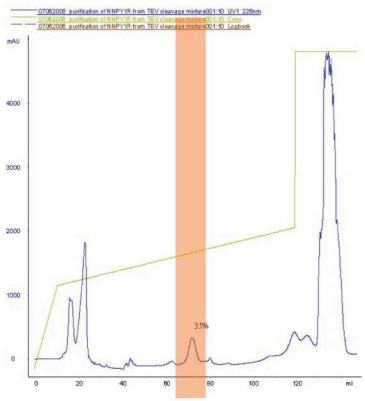
K - Chemical shift-mapping data for all neurohormones with N-Y1 (orange), N-Y2 (blue), and N-Y5 (green).



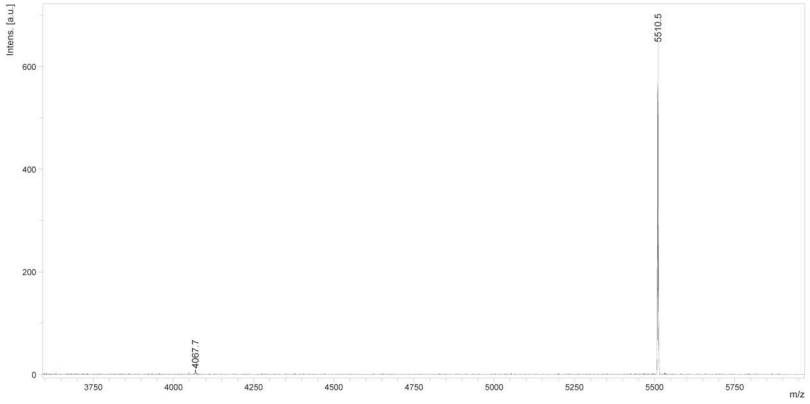
L - HPLC traces and MALDI-MS analyses of N-Y1, N-Y2, and N-Y5. The HPLC peaks corresponding to the peptides of interest are shaded in red.

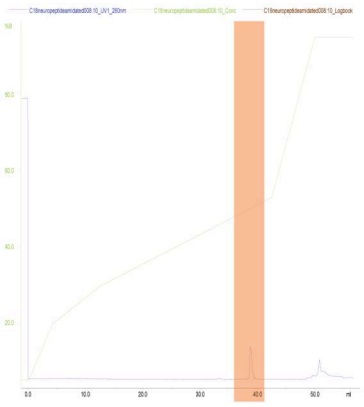
N-Y1



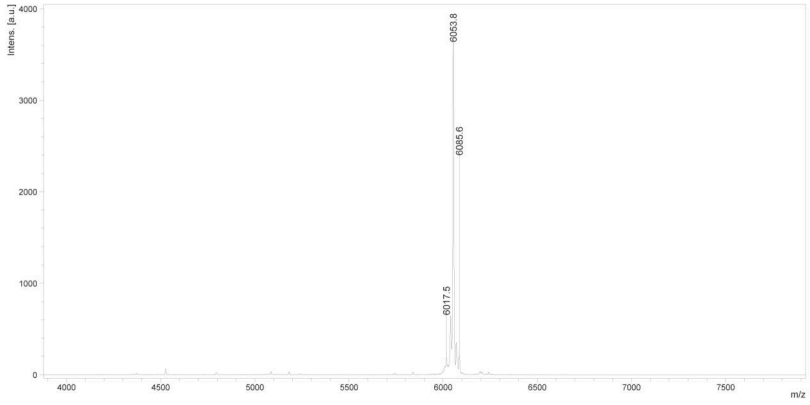


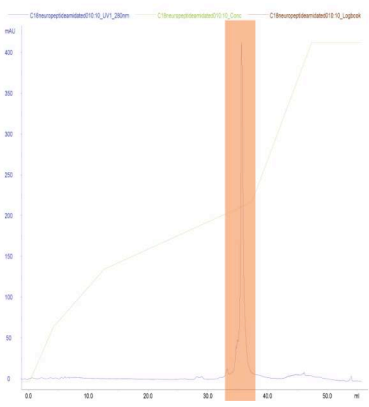
N-Y2





N-Y5





Chapter III: Grafting of extracellular loops of Y-receptors onto a soluble β-barrel scaffold

Abstract

In this chapter I present the attempt at creating a small, soluble model system that is capable of mimicking ligand-binding functions of large, membrane-embedded receptors. These receptors function in a modular way, in which a certain part of the molecule is responsible for a specific task. In our case one part is responsible for ligand binding, whereas other parts are involved in the signal transduction process or allosteric regulation, for instance. The full receptors are large and difficult to characterize structurally, but one can resort to studying confirmed or hypothesized modules thereof. I am trying to develop such an approach for studying one of the most important classes of cell surface receptors, the so-called G-protein coupled receptors (GPCRs). Cell surface receptors commonly are thought of consisting of three distinct modules or "domains": an extracellular, a transmembrane and an intracellular domain. We intend to study the extracellular part of a GPCR, which is responsible for binding the receptor's cognate ligand, by grafting it onto a stable supporting molecule, a so-called "scaffold". The resulting *minireceptor* should be easier to produce and characterize spectroscopically. GPCRs are of enormous clinical importance[1] and the availability of a suitable model for such receptors will be a valuable tool for the assessment of novel potential GPCR binding molecules.

1. Introduction

1.1 A new approach to structural information on membrane proteins

G-protein coupled receptors (GPCRs) are a class of membrane embedded proteins, spanning the hydrophobic membrane interior with seven transmembrane (TM) helices, connected by both three intracellular and extracellular loops[2]. GPCRs are of enormous clinical importance[1], and as long as no straight forward approaches to the three-dimensional structures of these molecules exist, the availability of models for such receptors will continue to be a valuable tool for the investigations of many different GPCRs. One sub-class of the GPCRs is believed to interact with their cognate ligands mainly through its surface exposed extracellular parts. Our work is inspired by the idea, that these

extracellular parts could be studied separately, i.e. in absence of the hydrophobic transmembrane helices, responsible for most of the problems arising in the work with GPCRs. Several approaches to realizing this idea have been formulated. The route we are following can be called “grafting approach” and can be briefly summarized as follows: If a small, soluble protein exists, which possesses a core determining its global fold and some surface exposed loops unimportant to that fold (i.e. a so-called “scaffold” protein), the external parts of a GPCR could be transferred to that stable core domain, without disturbing the overall global fold of the scaffold. The resulting chimeric protein could then be considered as a “minireceptor”, because it carries all the parts of the receptor hypothesized to be important for ligand binding (hence the term “receptor”), but at the same time shows the favorable characteristics of small, soluble proteins (hence the term “mini”). This idea is exemplified in figure 1.

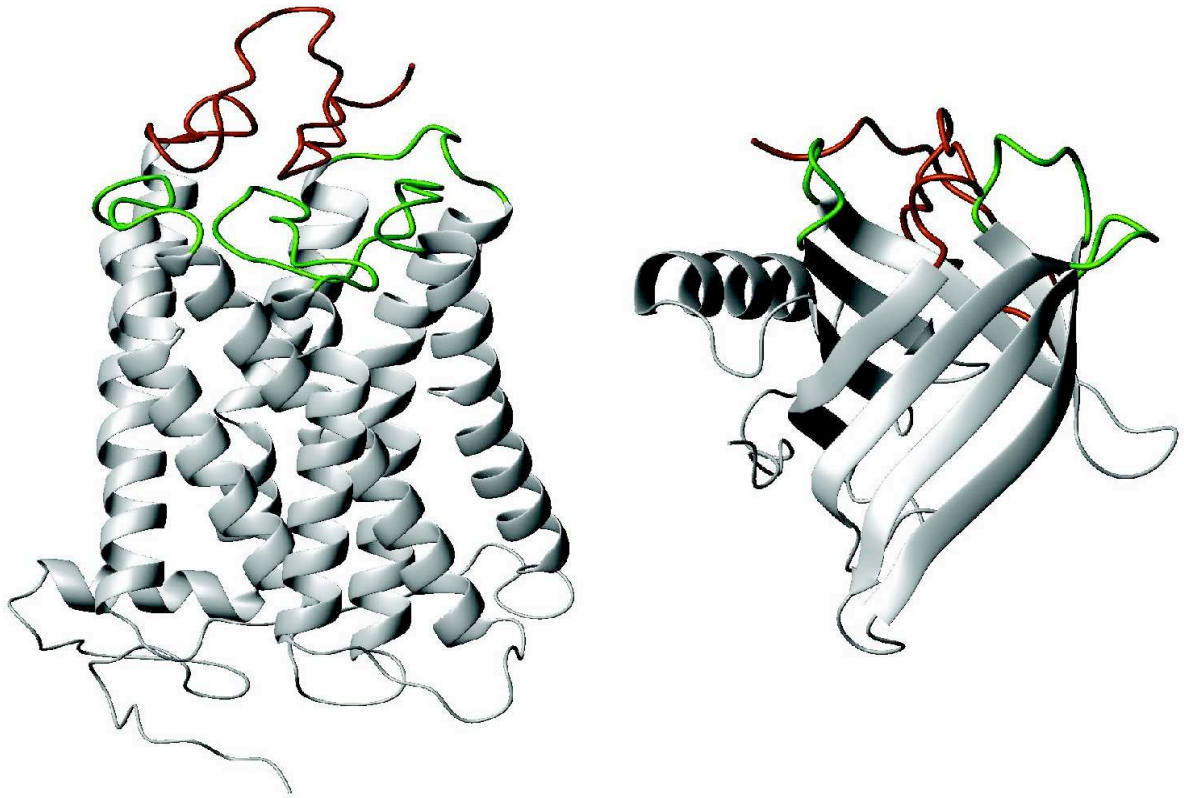


Figure 1: *Left: Structure of the prototypical GPCR bovine rhodopsin with extracellular loops and N-terminus color-coded in red and green, respectively. Right: Structure of a bacterial soluble β-barrel protein, which is suggested to be a suitable scaffold protein, with similar color-coding of loops and N-terminus.*

1.2 Scaffold molecules in nature and their use in synthetic protein chemistry

The advent of elaborate nucleotide synthesis and PCR schemes and highly efficient cloning techniques made it theoretically possible to produce large numbers of randomized derivatives of any given starting sequence. The theoretical number of different sequences of length n that can be composed of the 20 natural, proteinogenic amino acids rises with n in a factorial way. Therefore, already for relatively short sequences a manual expression and purification of each subsequence is practically not feasible. Through the invention of the "phage display" technology it became possible to screen large libraries for desired properties, thereby preventing the necessity to manually express, purify and test each potential candidate protein. The central idea of this technique is to create an as large as possible protein "library", where each protein and the genetic information that codes for it are physically linked. Provided an efficient selection strategy is present, one target protein along with its DNA-sequence can thus be pulled out of a mixture of billions of different proteins. Most commonly such a selection strategy consists of the surface immobilization of a source-protein and screening for target-proteins showing affinity to the source. With this technique the size of a "library" is no longer dictated by the number of sequences, which can be manually expressed and purified, but by other factors such as the transformation efficiency.

Such libraries of randomized sequences are very useful, and binders to a wide variety of different molecules can commonly be selected from one such given library. The mammalian immune system can be considered an elaborate *in vivo* protein library. The concept of generating libraries with structurally and functionally related compounds has, however, not evolved in mammals, but evolutionary much earlier in the class of the arachnids[3].

In vitro such libraries are generated by the application of randomizing PCR techniques. The bottle-neck for the generation of an as large as possible number of randomized sequences is the achievable transformation efficiency, i.e. how many colonies one can get from a given amount of randomized DNA. Still this renders the randomization of the complete sequence of a typical small protein not feasible. However this is also not desired. Instead, it is often more desirable to randomize only short surface-exposed sequences. The advantages of this approach are twofold: first, the number of theoretically possible variants is much smaller, and second the unchanged sequence can, in some instances, be considered to not undergo any structural changes. If the second criterion is fulfilled by a protein one speaks of it as a "scaffold". Such a scaffold is for instance formed in the immune system by the constant regions within the immunoglobulins or in the above mentioned arachnid peptide toxins by unchanged cysteine residues, which by forming intramolecular disulfide bond patterns, determine the

global fold of the peptides[3].

From the above-mentioned one can define certain characteristic features that a protein, which could possibly act as a scaffold, should present. First and foremost such a candidate scaffold should show high sequence diversity in conjunction with a conserved tertiary fold. In other words it should have been evolved by nature to be tolerant to changes in its amino acid sequence without losing its overall folding pattern. Additionally the protein family to which the potential scaffold belongs should possess a well-defined and structurally conserved hydrophobic core and a solvent accessible binding pocket, which is spatially and structurally well separated from the scaffold core. Ideally a potential scaffold candidate should already show diverse biochemical function among its family members. In other words nature should already have proven, that it can recognize ligands of great diversity.

1.3 Lipocalins as a scaffold

One protein family, which meets these requirements are the lipocalins. Lipocalins are a class of soluble β-barrel proteins, first described in 1985[4]. They are typically small (15-20 kDa), secreted proteins which often show the ability to bind small hydrophobic molecules[5-7]. The lipocalin fold is very well conserved and characterized by a short N-terminal 3_{10}-helix followed by an eight-stranded β-barrel and a C-terminal α-helix. The eight β-strands of the barrel (commonly referred to as strands A to H) are linked by short β-hairpin loops, except for the first loop (L1), which forms a large Ω-loop[7].

The lipocalin family possesses a well-conserved (hydrophobic) core, while the loop regions show more structural diversity[8,9]. Furthermore, many members of the lipocalin family possess a solvent-accessible site for the recognition of their ligand[10], which is spatially well separated from the core. Finally, the ligands recognized by lipocalins are of remarkable diversity[5]. Hence, they meet all of the above-mentioned requirements to be of use as scaffold molecules. Indeed lipocalins have been used successfully as scaffolds in numerous cases[8,11-13] and concomitantly robust expression and purification strategies for these proteins have been established. Under these premises we have selected the bacterial lipocalin Blc as our scaffold for the following reasons:

- the crystal structure of that protein has recently been solved[10], which enables rational replacements of its loops with the desired GPCR extracellular loops
- it carries three loops located on one and four on the other face of the barrel and thereby allows to introduce all three GPCR extracellular loops simultaneously; additionally the N-terminus of

Blc is mainly in an extended conformation and spans across both faces of the barrel, which allows the incorporation of the receptor N-terminus on either face of the barrel by attaching it an a suitable position within the extended Blc N-terminal sequence

- its sequence has been modified in order to allow for the expression as a monomeric, soluble protein in *E. coli*, yielding a protein in the molecular weight range amenable to high-resolution NMR without requiring expensive deuterium-labeling
- both periplasmic and cytoplasmic expression systems have been devised, so that disulfide bonds linking the loops or potential attachment points for reporter groups, such as spin labels, may be introduced
- it uses the *Strep*-tag II[14] technology for protein purification, which enables efficient purification and offers the possibility of immobilizing the protein on a streptavidin-coated surface for binding studies using SPR technology

It is commonly believed that the architecture of the helical bundle from different GPCRs is similar[15], and hence we extracted distances and relative orientations of the anchoring points (i.e. those points in the sequence, where the loops are anchored to the 7TM or the β-barrel scaffold) of all known GPCR structures and compared them to our designated scaffold Blc as shown in figure 2. Therein, β2-adrenergic receptor anchoring points are depicted in blue and are represented by the C_α atom closest to the extracellular membrane surface. The extracellular loops are schematically depicted as the arrows connecting two such anchoring points. Likewise the Blc anchoring points are represented in yellow.

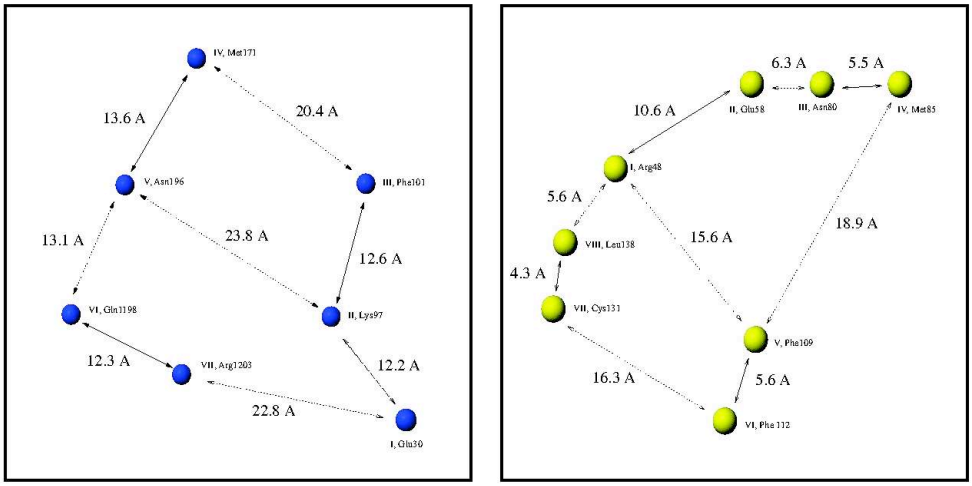


Figure 2: *Drawing displaying the relative orientations of loop-anchoring C_α atoms in the human β2-adrenergic receptor (β2-AR; left) and Blc (right). Each anchoring point is labeled with the number of the transmembrane α-helix (β2-AR) or β-strand (Blc) to which it corresponds (roman numbers).*

From the comparison of the distances between the hypothesized anchoring points of the Y-receptors and the confirmed anchoring points for the Blc scaffold, it is obvious that they do not match exactly. Therefore, it is to be expected that any grafted loops will probably require the simultaneous co-transfer of some flexible linker residues in order to relieve structural constraints, which would be imposed on the loops from the scaffold otherwise. Glycine- and serine-glycine-rich repeats have been found to be suitable linkers in studies having encountered similar such problems previously. However, the figure also reveals, that the relative orientation of the loop anchor points corresponds rather well between the model GPCR rhodopsin and our target scaffold Blc. Considering that it is unclear, whether loops of the Y-receptors are well-structured at all, and taken into account that i) Boivin *et al.*[16] could detect interactions described by K_ds as low as $1.4 \bullet 10^{-6}$M between immobilized biotinylated peptides corresponding to the second extracellular loop (ECL2) of the urotensin receptor and its natural ligands and ii) the fact that we could demonstrate binding of the peptides to the free N-terminal domain, and iii) the possibility to from a disulfide bond between the 1st and 2nd loop to further restrain the system in a correct manner indicates, that a system capable of tight binding to the neurohormones may be successfully designed.

1.4 Structural characterization of the lipocalins

Lipocalins have a very conserved characteristic fold. Its core is an eight-stranded β-barrel. N-terminally the barrel is preceded by a short 3_{10}-helix and C-terminally it is followed by a regular α-helix[6,17]. The secondary structure and its characteristic packing are depicted schematically in figure 3.

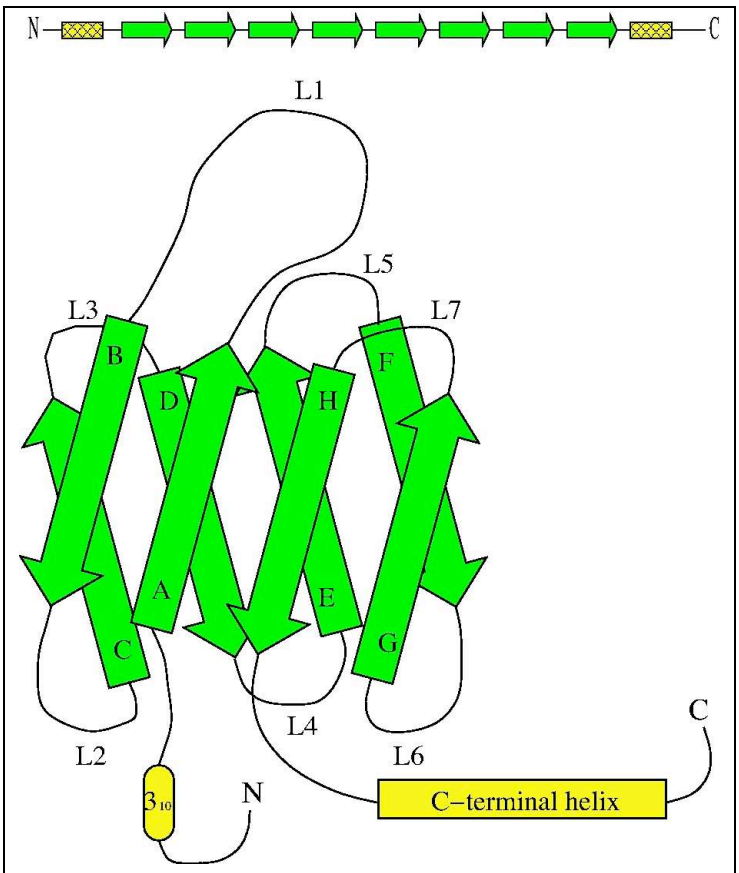


Figure 3: *The extended secondary structure elements of a typical lipocalin are shown in an extended representation in the upper picture, while their schematic packing against each other is shown in the lower picture. Green arrows represent the β-strands, while the yellow ellipsoid and the yellow rectangle represent the N-terminal 3_{10}-helix and the C-terminal α-helix, respectively.*

The bacterial lipocalin for which the crystal structure was determined[10] also shows the above depicted core secondary structure. Additionally it carries a short two-stranded β-sheet at its N-terminus. The residues of the N-terminal β-strand, however, are an artifact from the chosen cloning strategy. The relevance to the overall structure of Blc of this β-sheet is thus disputable. The secondary structural elements mapped onto the amino acid sequence of the construct used in the crystallographic study are shown in figure 4. Additionally this figure shows a sequence comparison between the construct used in the crystallographic study, the wt-Blc and the sequence of our construct. Our construct lacks the first 19 residues of the construct used in the crystal structure study and carries an additional 9 residues at the C-terminus, which serve as an affinity tag, but is almost identical otherwise. Most notably the single cysteine residue present at position 131 in the crystal structure is replaced by a serine (figure 4).

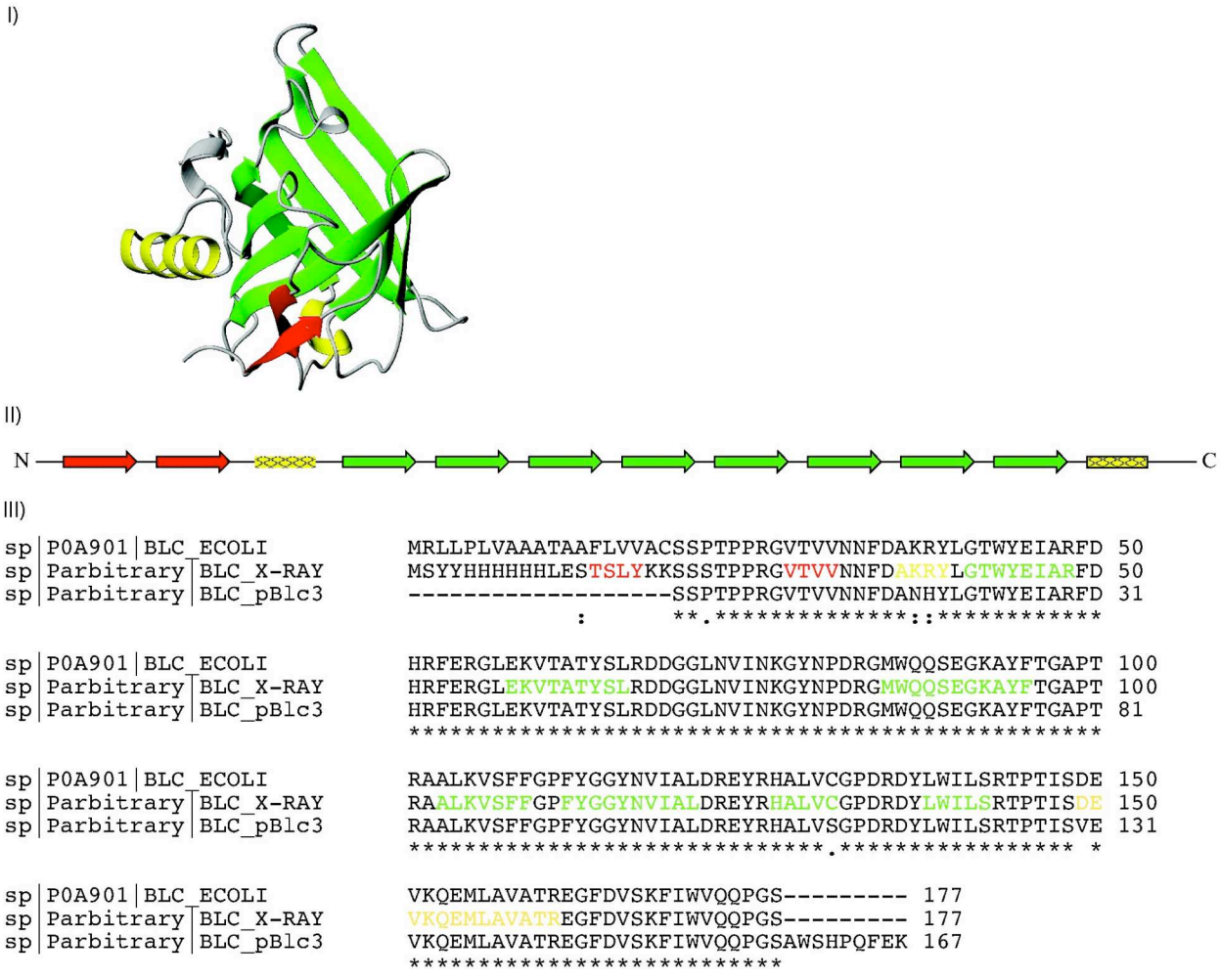


Figure 4: *I) Crystal structure of Blc (1QWD.pdb) with β-strands of the barrel colored in green, N-terminal 3_{10}- and C-terminal α-helix colored yellow and short antiparallel β-sheet in the N-terminus colored red. II) The extended secondary structure elements of Blc according to the crystal structure. Color-coding as in I). III) Shows a sequence alignment produced with CLUSTALW of the wildtype Blc sequence as deposited in the SwissProt database (P0A901 BLC_ECOLI), the sequence of the X-ray crystal construct (BLC_X-RAY) and the Blc construct used by us (BLC_pBlc3).*

1.5 The receptors of the NPY-family neuropeptides: The Y-receptors

Our group has studied a wide variety of neurohormones of the neuropeptide Y (NPY) family using high-resolution NMR[18-21] and has proposed a model for the binding of these hormones to their cognate receptors that includes membrane-association as an initial step[20]. Structural and mechanistic insight into such binding processes is anticipated to provide a basis both for the rational design of novel, and the improvement of already existing receptor-agonists and antagonists. Because of the detailed studies that have been conducted in our group on the above mentioned neurohormones and the detailed

biophysical and biochemical characterization in other groups[22-25], the receptors for the NPY-family neuropeptides are called Y-receptors and have been chosen as a model system for the investigation of ligand binding to GPCRs.

The neuropeptides exert their effects via several receptor subtypes called Y-receptors. Four main receptors, named Y1-[26-28], Y2-[29], Y4-[30,31] and Y5-[32], have been cloned so far. All identified Y-receptors act via pertussis toxin-sensitive G proteins of the G_i family. They range in size from 375 to 455 residues and show the prototypical characteristics of GPCRs of subfamily 1b. The different receptor subtypes are localized in various tissues, both in the central nervous system and in the periphery. NPY and peptide YY (PYY) bind equally well to the receptors Y1, Y2, and Y5 (nanomolar to sub-nanomolar dissociation constants). Only the pancreatic polypeptide (PP) shows selectivity towards the Y4-receptor (picomolar dissociation constant). The ability of NPY and PYY to bind to three different receptor subtypes may be related to their conformational flexibility, which enables the peptides to adopt more than one energetically favorable conformation.

The table below gives a summary about the most important data concerning the four major Y-receptors:

Receptor	**Y1**	**Y2**	**Y4**	**Y5**
K_i NPY [nM]	0.81[a]	0.02[a]	1.9[a]	0.19[a]
K_i PYY [nM]	1.1[a]	0.01[a]	1.1[a]	0.06[a]
K_i PP [nM]	>100[a, b]	>1000[a, b]	0.04[a, b]	27[a, b]
amino acids	384	381	375	455
major occurrence	periphery[c], hypothalamus[d]	CNS[e, f, g]	intestine, pancreas[h]	hypothalamus[i]
related action	vasoconstriction, anxiolysis[c, d, j]	memory, epilepsy, secretion[c, k, l]	gastro-intestinal regulation[m]	food intake[i]

Table 1: *Biochemical, biophysical and physiological characteristics of the four major families of the Y-receptors. a) McCrea, Regul. Pept., **87**, 47-58, 2000; b) Small, Proc. Natl. Acad. Sci. U S A, **94**, 11686-91, 1997; c) Wahlestedt, Med. Biol., **64**, 85-8, 1986; d) Wahlestedt, Science, **259**, 528-31, 1993; e) Gehlert, Mol. Pharmacol., **49**, 224-8, 1996; f) Gerald, J. Biol. Chem., **270**, 26758-61, 1995; g) Rose, J. Biol. Chem., **270**, 22661-4, 1995; h) Lundell, J. Biol. Chem., **270**, 29123-8, 1995; i) Gerald, Nature, **382**, 168-71, 1996; j) Grundemar, Br. J. Pharmacol., **105**, 45-50, 1992; k) Flood, Peptides, **10**, 963-6, 1989; l) Potter, Regul. Pept., **25**, 167-77, 1989; m) Schwartz, Gastroenterology, **85**, 1411-25, 1983*

Sequence comparison shows, that among these four receptors, the Y1 and Y4 are the most closely related (42% sequence homology) while Y2 and Y5 are equally distant from each other and from the Y1/Y4 pair[33,34]. These relationships are also reflected when the predicted soluble N-termini of the four receptors are compared, as is shown in the sequence alignments below.

```
NPY1R_HUMAN      MN-STLFSQVENHSVHS-NFSEKNAQLLAFENDDCHLPLAMI 40
NPY4R_HUMAN      MNTSHLLALLLPKSPQGENRSKPLGTPYNFS-EHCQDSVDVM 41
                 ** * *:: :  :* :. * *:  .   *. :.*: .: ::

NPY2R_HUMAN      MGPIG-AEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELID-STKLIEVQ 50
NPY5R_HUMAN      MSFYSKQDYNMDLELDEYYNKTLATENNTAATRNSD-FPVWDDYKSSVDDLQ 51
                 *.  .  : : :  ::*   :  ..:... .  .*  *  * .:.: ::*
```

In several studies the N-termini of Y-receptors have – as typical for GPCRs of the subfamily 1b - been shown to be involved in ligand binding[35,36]. We therefore sought to express the 40 to 50 residue N-terminal fragments of the four Y-receptors and characterize them by NMR.

Their ligands, the members of the NPY family, are C-terminally amidated polypeptides comprised of 36 amino acids. Their postulated role in the regulation of food-uptake[37,38] has stirred vigorous research in many pharmaceutical companies. Seven positions are absolutely conserved among all species of NPY, PYY and PP. These are Pro5, Pro8, Gly9, Ala12, Tyr27, Arg33 and Arg35[39] (figure 5). The first atomic resolution structure of a member of the NPY family was obtained for avian PP (aPP) by X-ray crystallography[40] and later for bovine PP (bPP) in solution by NMR[41]. Residues 1 to 8 form a type-II polyproline helix, which is followed by a type-I β-turn connecting to residues 15 to 32, which form an α-helix, and the four most carboxy-terminal residues are in a flexible loop conformation (figure 5). A surprisingly stable helical hairpin is formed by backfolding of the polyproline helix onto the α-helix. This structural motif is commonly referred to as the PP-fold. The tertiary contacts, which stabilize this fold are formed by interdigitation of the conserved proline residues in position 2, 5, and 8 on the polyproline helix and the tyrosines 20 and 27 on the α-helix[42]. The solution structure of PYY in aqueous solution was shown to be highly similar[21,43]. Surprisingly, in the highly homologous NPY the backfolding is absent[44] (figure 5). All three members of the NPY family have also been characterized in presence of dodecylphosphocholine (DPC) micelles mimicking a biological membrane (figure 5). Whereas in aqueous solution PYY and PP are structurally similar, in the presence of micelles PYY[21] and NPY[45] adopt a similar conformation, with PP taking an alternative conformation[19]. It was found that pNPY and pPYY interact via the hydrophobic face of their C-terminal α-helix with the micelle, while the N-terminus freely diffuses in solution. bPP also interacts with the micelle via its C-terminal α-helix, but the N-terminus is also loosely associated with the micelle surface. It has been proposed, that this association is mediated by Tyr7 in bPP[21], which has a favorable free energy of transport into the water-membrane interface[46], in contrast to the Asn7 and Ala7 found in pNPY and pPYY, respectively. In addition also the C-terminal pentapeptide, which contains particularly important residues for receptor binding[23], differs structurally in bPP from the pNPY/pPYY pair[21].

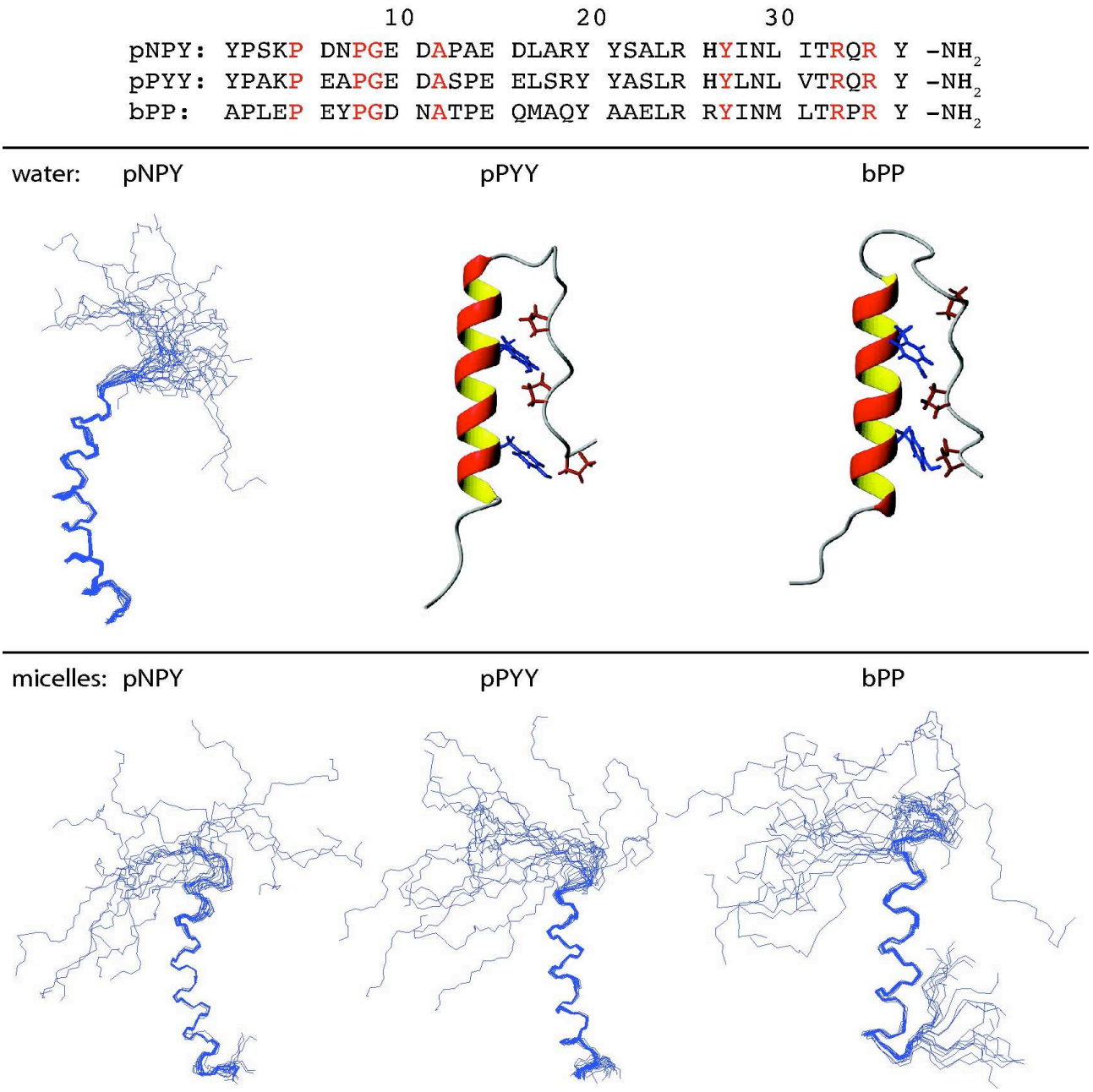


Figure 5: *Neurohormones of the NPY family. Top: Sequence comparison of porcine NPY (pNPY) and PYY (pPYY) with bovine PP (bPP). Residues, which are conserved among all three neurohormones throughout all species, are colored red. Bottom: Threedimensional structures of pNPY, pPYY, and bPP in water and in DPC micelles. In the structures of pPYY and bPP in water, the interdigitating Pro and Tyr residues are colored red and blue, respectively.*

Little is known about the structure of any of the Y-receptor subtypes, nor have high-resolution experimental data on the complex formed between these and the neurohormones been published. Nevertheless, in many photoaffinity-labeling studies the contact points between receptors and their ligands have been assigned to the extacellular face of the GPCR. In case of human NPY at the human Y1-receptor Walker *et al*. have postulated a prominent role for Asp residues of the Y1-receptor in the ligand binding process[25]. In particular, an Asp residue at the interface between the 6th TM domain and the third extracellular loop is conserved in all known Y-receptor sequences. Other potential sites of interaction were postulated to reside within the N-terminal domain and the first extracellular loop[24].

2. Results

2.1 Expression of N-terminal domains in isotopically labeled form

2.1.1 Expression and purification of NY2R

We first sought to find out, if some structural entities in the Y-receptors exist, which show some basic affinity towards the neuropeptides, even when removed from their natural context of the 7TM Y-receptor. To this end we produced the N-terminal fragment of the Y2-receptor (NY2R) both in unlabeled and in ^{15}N-labeled form.

Isotope labeling of the investigated peptides was required for the study of backbone dynamics using ^{15}N relaxation and for chemical shift mapping experiments for the study of macromolecular interactions. Such labeling precludes the usage of peptides produced from solid phase synthesis and necessitates recombinant production. For reasons of simplicity *E. coli* is generally the preferred expression host. To avoid rapid degradation in *E. coli*, the peptide needs to be linked to a (more) stable fusion partner. Specific cleavage from the fusion partner can be accomplished for systems for which a specific hydrolase (*e.g.* a ubiquitin hydrolase, TEV hydrolase) is available or by introducing a unique cleavage site, either a protease-sensitive site or a site prone to chemical cleavage, for example by CNBr[47] or hydroxylamine[48]. In our case CNBr cleavage was incompatible with the presence of Met residues in the NY2R sequence, and poor efficiency was observed with hydroxylamine. Therefore enzymatic cleavage had to be used.

Since the Y2-receptor N-terminal fragment should be reasonably water-soluble, we initially decided to express it as a C-terminal fusion to N-terminally decahistidine-tagged yeast ubiquitin. After purification of the fusion construct by Ni-affinity chromatography, the desired peptide was liberated through treatment with yeast ubiquitin hydrolase (YUH). This system allowed the recovery of about 6 mg of ^{15}N-labeled NY2R from 1 L of culture. In figure 6 the purification procedure from overexpression of the fusion construct in *E. coli* to the final, purified NY2R is shown.

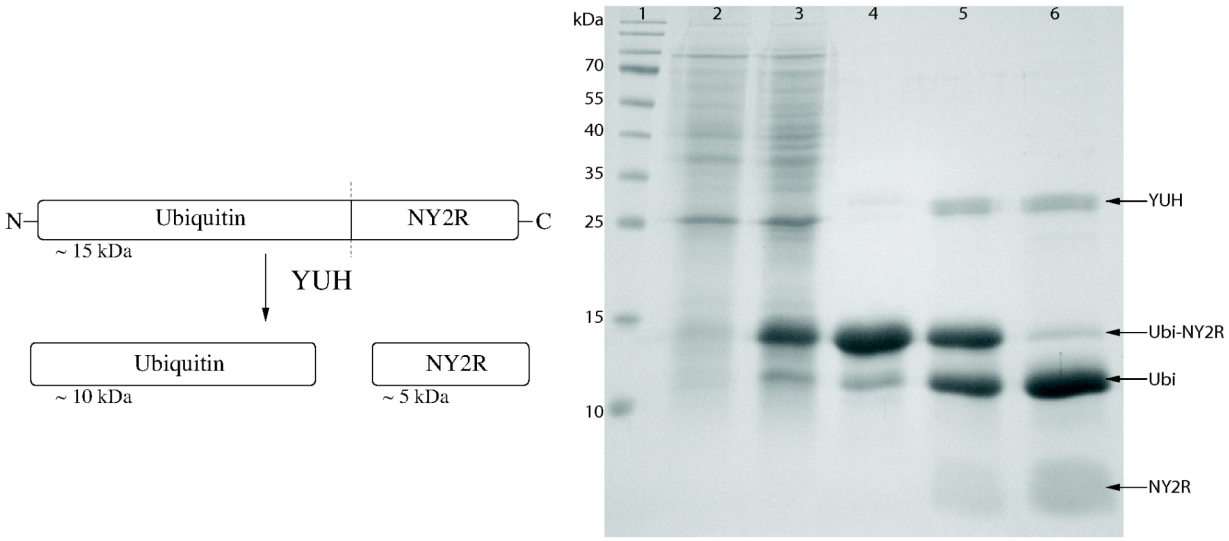


Figure 6: *Left: Schematic representation of the ubiquitin fusion system for the recombinant production of NY2R. Right: SDS-PAGE of the expression, purification and cleavage of the ubiquitin-NY2R fusion construct. Lane 1: molecular weight marker; lane 2: cells expressing ubiquitin-NY2R before induction; lane 3: cells expressing ubiquitin-NY2R after induction; lane 4: ubiquitin-NY2R after purification by Ni-NTA chromatography; lane 5: cleavage reaction mix at the beginning of the reaction; lane 6: cleavage reaction mix at the end of the reaction.*

The N-termini of the other three Y-receptors were produced by other members in our group as reported in Zou *et al.*[49] and in chapter 1 of this thesis. A comparison between the N-terminal sequences of the four Y-receptors is shown in figure 7. It can be seen that the NY1R and the NY4R and the NY2R and NY5R share pair wise similarity among each other.

```
NPY1R_HUMAN      MN-STLFSQVENHSVHS-NFSEKNAQLLAFENDDCHLPLAMI 40
NPY4R_HUMAN      MNTSHLLALLLPKSPQGENRSKPLGTPYNFS-EHCQDSVDVM 41
                 ** * *:: :  :* :. * *:  .   *. :.*: .: ::

NPY2R_HUMAN      MGPIG-AEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELID-STKLIEVQ 50
NPY5R_HUMAN      MSFYSKQDYNMDLELDEYYNKTLATENNTAATRNSD-FPVWDDYKSSVDDLQ 51
                 *.  .  : : :  ::*   :  ..:... .  .*  *  * .:.: ::*
```

Figure 7: *Sequence alignments of the human Y-receptor N-terminal domains.*

2.1.2 Structural characterization of NY2R by NMR and interaction studies with neurohormones

From preliminary shift mapping experiments it was clear that the neuropeptides and the NY2R interacted in a specific manner with each other. This was mainly evident for NPY from two

complementary sets of titration experiments:

The changes in the ^{15}N-labeled neuropeptide NPY proton-nitrogen correlation spectra ([^{15}N,^{1}H]-HSQC) upon addition of unlabeled NY2R (U-NY2R) could be attributed to specific residues within the sequence of the neuropeptides, because for these the complete sequential assignment was known from previous work. A comparison of spectra of all three members of the NPY family before and after addition of 2 equivalents of NY2R is shown in figure 8. The observed changes are much more prominent in NPY than in PYY and PP. For NPY the most pronounced changes can be observed for residues A14, E15, Y21 and H25.

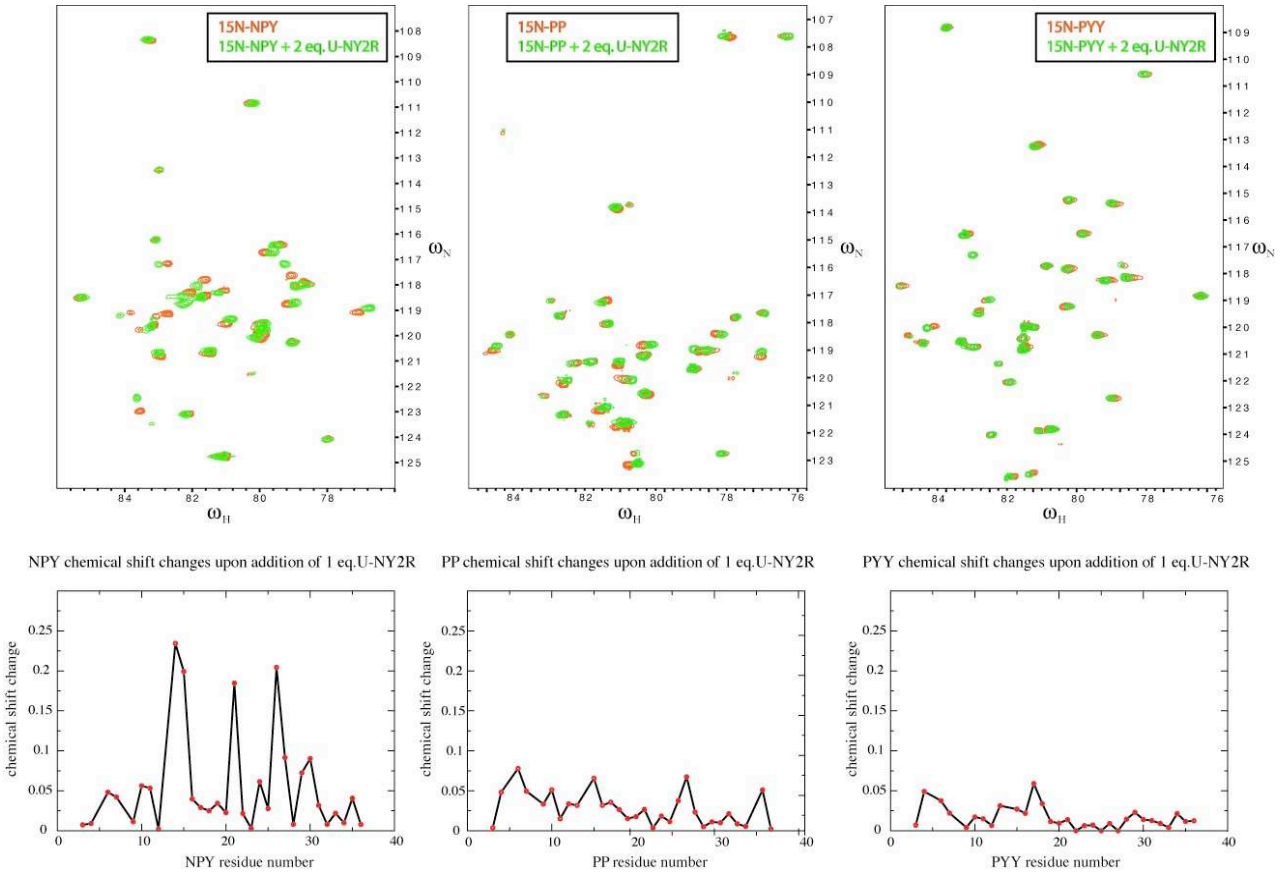


Figure 8: *The upper row shows [^{15}N,^{1}H]-HSQC spectra of 0.2 mM ^{15}N-labeled neuropeptides in 20 mM d_{13}-MES pH 6, 300 mM d_{38}-dodecylphosphocholine (DPC) at 37 °C without (red) and with 2 equivalents of unlabeled NY2R (U-NY2R) added. The lower row shows the chemical shift changes al along the sequence of the respective neuropeptide.*

In order to be able to perform the complementary type of analysis for the NY2R upon titration with the unlabeled neuropeptides, it was first necessary to assign the resonances of NY2R peptide. Sequence-specific resonance assignments were obtained using the strategy developed by Wüthrich and coworkers[50]. Due to extensive resonance overlap of the poorly folded peptide ^{15}N-resolved three-

dimensional TOCSY and NOESY data had to be utilized for this task. Below some resolved spin systems from the TOCSY and some inter-residual connectivities from the NOESY are shown.

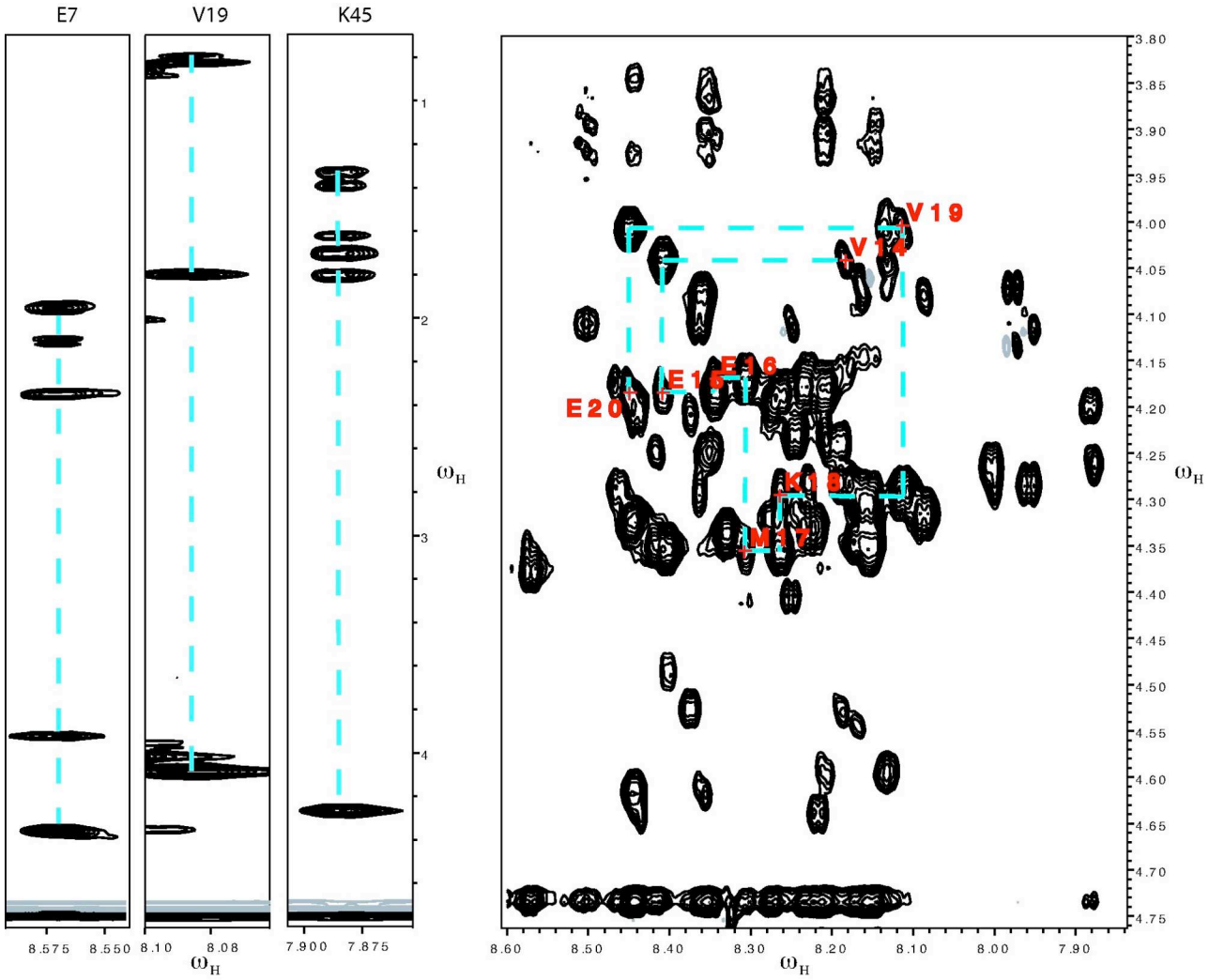


Figure 9: *TOCSY strips and NOESY of 0.9 mM NY2R, 300 mM d_{38}-DPC, 20 mM d_{13}-MES pH 5.5, 10% D_2O at 37 °C. On the left five the amide region TOCSY strips of some selected, well resolved spin systems are shown. On the left the H^N-H^α region of the NOESY spectrum is shown. The connectivity between seven spin systems is highlighted.*

A table with all the amide proton chemical shifts of NY2R is given in Appendix D.

This resulted in a complete assignment of the 50 residue peptide. In figure 10 a [^{15}N,^{1}H]-HSQC spectrum is shown with the peaks labeled with the corresponding amino acid one letter code and number. A table with all the amide proton chemical shifts of NY2R is given in Appendix D. In order to test for interaction between NY2R and NPY a titration experiment was carried out. A sample containing 0.1 mM ^{15}N-NY2R, 300 mM d_{38}-DPC, 20 mM d_{13}-MES pH 5.5 was prepared. [^{15}N,^{1}H]-HSQC spectra of this sample were measured at 37 °C after the addition of 0, 1, 2, 4 and 10 equivalents

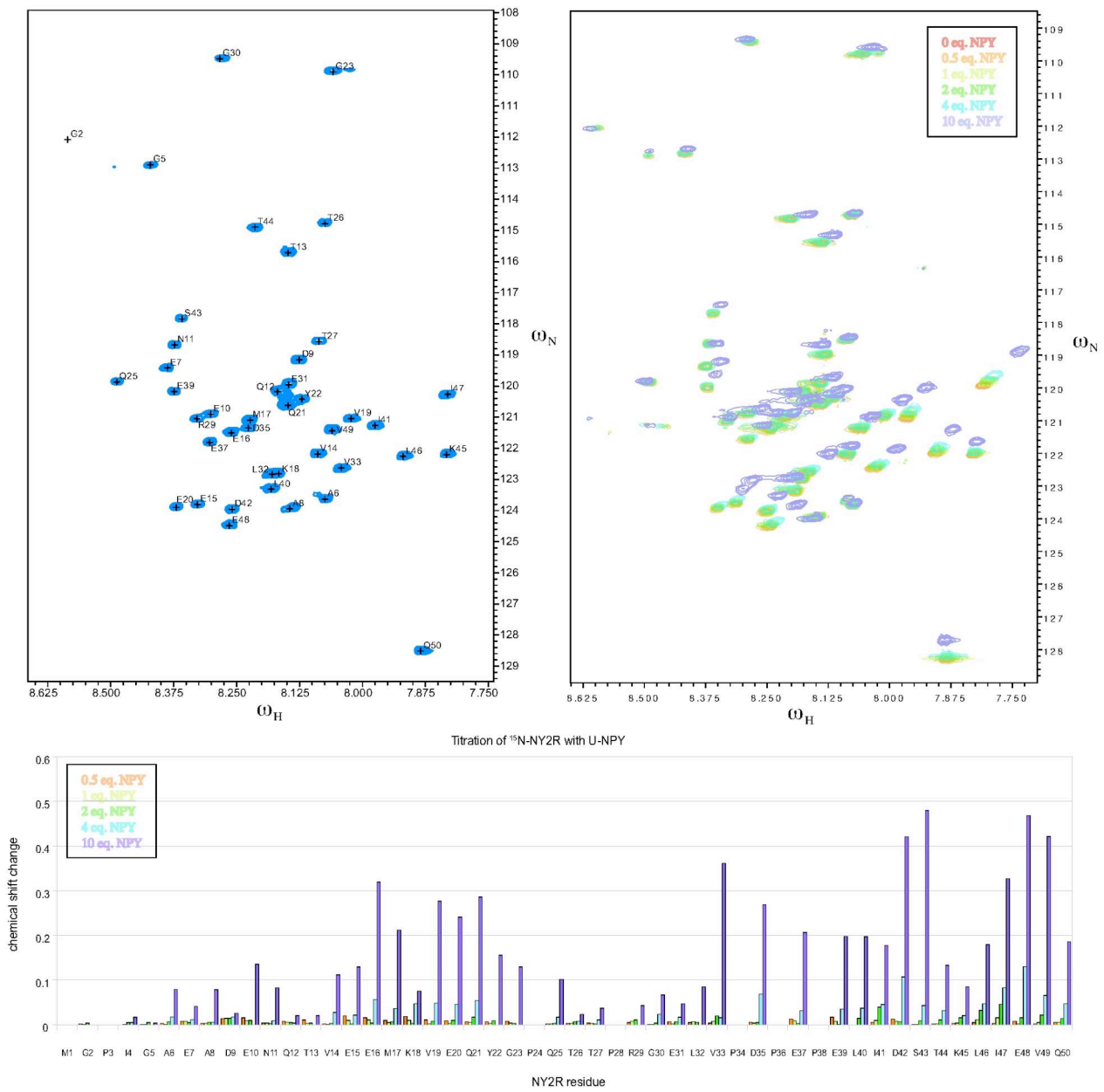


Figure 10: *Top left: [^{15}N,^{1}H]-HSQC spectrum of 0.9 mM N-NY2R, 300 mM d_{38}-DPC, 20 mM d_{13}-MES pH 5.5, 10% D_2O at 37 °C. Each peak is assigned as shown by the one letter amino acid code and its number in the sequence. Top right: Overlaid [^{15}N,^{1}H]-HSQC spectra of 0.1 mM ^{15}N-NY2R, 300 mM d_{38}-DPC, 20 mM d_{13}-MES pH 5.5 after the addition of 0 (red), 0.5 (orange), 1 (yellow), 2 (green), 4 (cyan) and 10 (purple) equivalents of unlabeled NPY. Bottom: Chemical shift changes in the spectra of ^{15}N-NY2R after the addition of 0.5, 1, 2, 4 and 10 equivalents of unlabeled NPY.*

of unlabeled NPY. A graphical representation of the observed chemical shifts is given in figure 10. The most pronounced chemical shift changes are observed for a central and a C-terminal fragment of NY2R both rich in negatively charged amino acids.

Interactions of two peptides in solution should, in principle, not only be evident from chemical shift changes, but also through changes in the local dynamics in the two interacting molecules. A way of estimating peptide dynamics is the measurement of the heteronuclear ^{15}N{^{1}H}-NOE. This kind of

measurement was carried out for ^{15}N-NY2R in the absence and in presence of unlabeled NPY. The observed values are depicted in figure 11. As mentioned the most prominent chemical shift changes were observed for a central (residues 14-22) and a C-terminal (residues 33-50) fragment. This is more or less in accordance with the changes observed in the heteronuclear NOE of NY2R upon addition of NPY. Even though the changes are small, they are centered roughly around the same two regions.

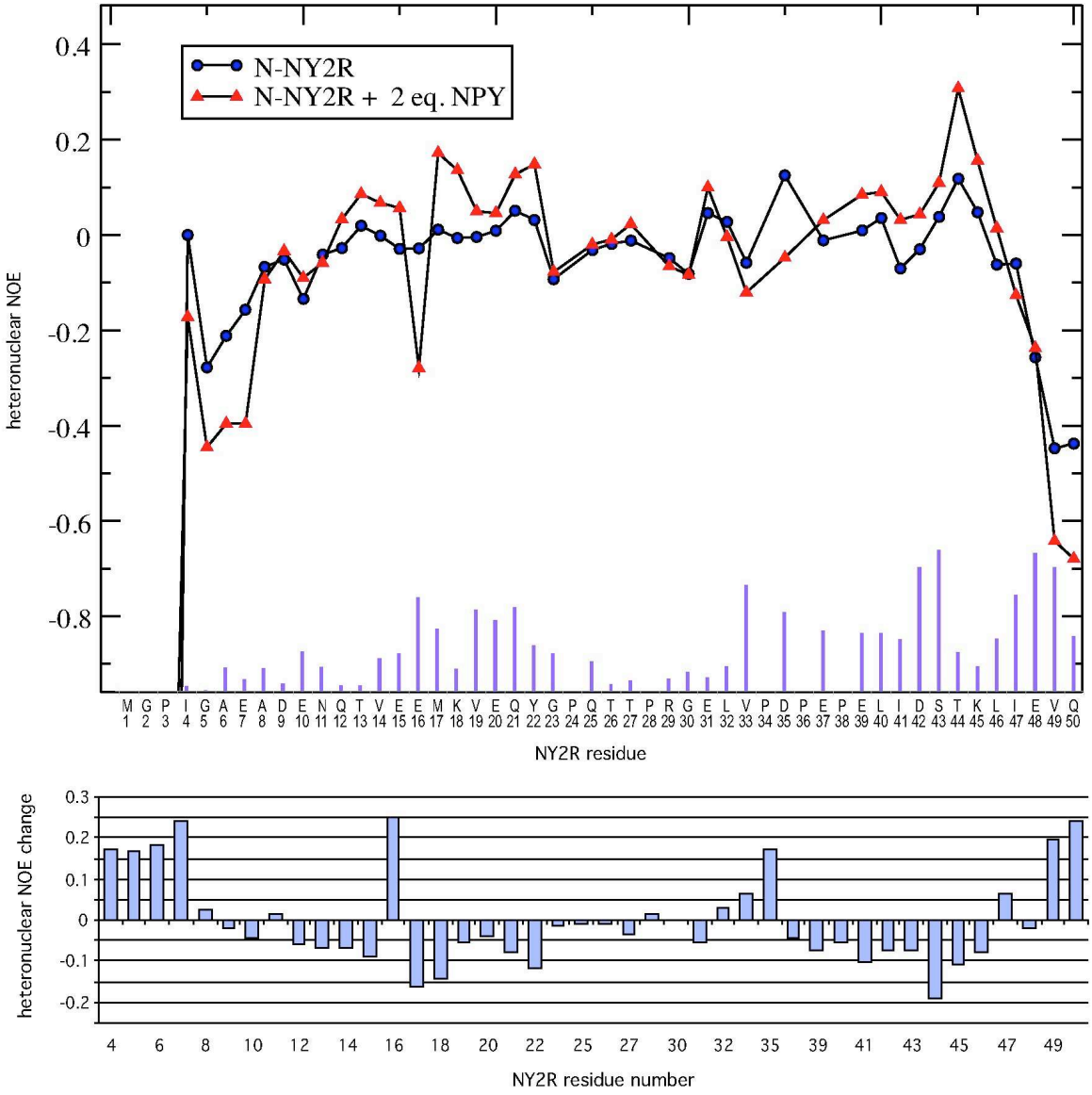


Figure 11: *Top: $^{15}N\{^{1}H\}$-NOE of 0.9 mM N-NY2R in 20 mM d_{13}-MES, 300 mM d_{38}-DPC at 27 °C without (blue circles) and with 2 equivalents of unlabeled NPY (red triangles) added. Bottom: Difference of the $^{15}N\{^{1}H\}$-NOE between the sample with and without 2 equivalents of unlabeled NPY.*

As both the chemical shift changes and the heteronuclear NOE changes upon addition of NPY show a similar behavior, with two distinctly influenced regions around Met17 and Thr44, the N-terminus of the Y2-receptor might be considered a good candidate for the transfer onto our β-barrel scaffold as some interaction with the neurohormones seems to be present also outside the context of its natural 7TM receptor. The fact, that segments rich in negatively charged residues are displaying the largest interactions, and the weakness of the interaction indicates that the N termini display weak electrostatic interactions to the neurohormones, a feature that has been investigated in more detail for the PP-NY4 interaction by Zou *et al.*[51].

2.2 Characterization of the soluble scaffold

2.2.1 Expression and Purification of Blc

For later grafting experiments it would be beneficial to have the scaffold thoroughly characterized spectroscopically possible. Ideally a complete sequential assignment of at least the backbone should be present to be able to easily distinguish resonances from grafted sequences from scaffold resonances in later experiments. Also, the knowledge about the dynamics of the scaffold molecule might be beneficial to choosing good points in the sequence to attach or insert the foreign Y-receptor protein sequences.

We selected the bacterial lipocalin (Blc) as a possible soluble scaffold. This gene has been cloned into the pASK-IBA5 and pASK-IBA3 vectors in the group of Prof. Skerra giving us the possibility to express the Blc both in the cytoplasm (pBlc5) and in the periplasm (pBlc3). The insert is composed of 501 bp coding for the 167 residues of Blc plus, in the case of pBlc3 a sequence coding for a bacterial leader peptide, which directs the mature protein into the periplasm and is cleaved off after membrane translocation by the *E. coli* signal peptidase I. In our case the leader peptide is the OmpA signal peptide[52]. The periplasmic expression of proteins offers several advantages: (i) the recombinant protein can form disulfide bonds (if present) in the oxidative periplasmic space, (ii) it is separated from the mostly cytosolic host cell proteases, (iii) the outer membrane of gram-negative bacteria such as *E. coli* can be selectively disrupted under mild conditions without disrupting the cell's inner membrane, avoiding "contamination" of the recombinant protein with cytosolic proteins and (iv) the periplasmic space is open for diffusion of solutes up to 600 Da, allowing the addition of folding enhancing substances to the culture medium[53,54]. A disadvantage of periplasmic expression of recombinant proteins, however, is that it is often accompanied by aggregate formation. Advantageous to the

cytosolic expression of recombinant proteins is the fact that it often results in higher yields. For strong overexpression, however, also in the cytosol the formation of insoluble aggregates, so-called "inclusion bodies", is often observed. The yield of correctly folded, soluble protein can in both cases be increased by lowering the growth temperature[55]. This however, comes at the cost of a slower growth rate.

The pASK-IBA vectors, which are commercially available from IBA (http://www.iba-go.com), are derived from the plasmid pASK75[56]. Their expression cassette is under transcriptional control of the tightly regulated *tetA* promoter/operator and is flanked by *XbaI* and *HindIII* restriction sites. The *tet* repressor is encoded on the same vector and is under constitutive expression of the β-lactamase promoter. This ensures a tight repression in the absence of inducer[14,57,58] and a stoichiometric relationship between cellular repressor molecules and the plasmid copy number. The strength of the *tetA* promoter is comparable to the conventionally used *lacUV5* promoter. It can be induced by the addition of anhydrotetracyclin at a concentration, which is not antibiotically effective.

Just downstream of the *XbaI* restriction site the pASK-IBA3 vector contains the OmpA signal sequence, which mediates secretion of the target protein into the periplasmic space. In the pASK-IBA5 vector this sequence is missing, resulting in cytoplasmic accumulation of the target protein. Both parent vectors contain just upstream of the *HindIII* restriction site a sequence coding for the *Strep*-tag II104. This tag is derived from the *Strep*-tag[59], which is itself derived from the biotin binding epitope of streptavidin. Both these tags contain a HPQ motif [60], which is characteristic for streptavidin binding peptides and has the property of binding to streptavidin competitively with biotin, which is streptavidin's natural ligand. This allows for the use of very gentle conditions for the elution of bound *Strep*-tag fusion proteins from streptavidin affinity columns with dilute biotin solutions[59]. The *Strep*-tag II is composed of the sequence Asn-Trp-Ser-**His-Pro-Gln**-Phe-Glu-Lys (NWS**HPQ**FEK) and was developed from the *Strep*-tag mainly as a variant more suitable for all N-terminal, C-terminal and also internal fusion to target proteins[61]. Targeted random mutagenesis of a flexible loop in streptavidin, mutants could be selected with enhanced affinity both for the *Strep*-tag and for the *Strep*-tag II[62]. One such mutant showed markedly improved performance in the purification of *Strep*-tag II fusion proteins, when coupled to a chromatographic support[62]. This mutant was later named StrepTactin and is commercially available coupled to different resins for chromatography from IBA.

Good expression yields for Blc have been obtained in the bacterial strain JM83. However, as this is a proline-auxotrophic strain it is not ideally suited for the growth on minimal media, which are commonly employed in bio-NMR to obtain isotopically labeled proteins at reasonable costs. We therefore decided to evaluate the standard BL21 (DE3) and C41 (DE3) [63], strains for their efficiency in expressing the two different Blc constructs. Cell pellets from LB-cultures were analyzed by SDS-

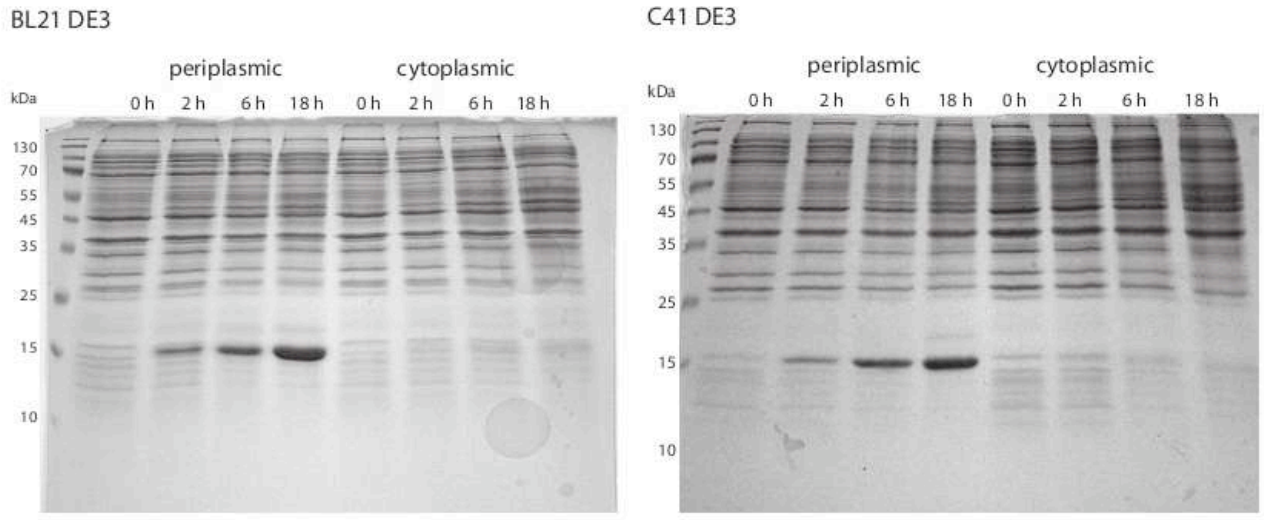


Figure 12: *Test expressions of the periplasmic and cytoplasmic Blc constructs in LB-broth. Two cell strains, BL21 (DE3) and C41 (DE3), were tested. Samples of the uninduced cultures and after 2, 6, and 18 hours after induction are shown. The periplasmic induction seems to work equally well for both cell strains, while the cytoplasmic expression failed in both strains under the chosen conditions.*

PAGE for expression efficiency after induction. Figure 12 shows the induction profiles of BL21 (DE3) and C41 (DE3) harboring the plasmids pBlc3 and pBlc5, respectively.

Based on the expression profile shown in the above gel, BL21 (DE3) seemed to be slightly more effective for the expression of the periplasmic Blc construct, than the C41 (DE3) strain. The cytoplasmic expression failed in both strains under the chosen conditions. We therefore selected the periplasmic expression system in BL21 (DE3) cells for the further production of the Blc.

2.2.2 NMR spectroscopic characterization of Blc

To assess the expression and purification efficiency a 1 liter culture of BL21 (DE3) pBlc3 in unlabeled minimal medium was grown. Blc was purified according to the procedure described in the materials and methods part to yield approximately 2 ml of a 1 mM unlabeled Blc (U-Blc) solution. A 1D ^{1}H-spectrum was recorded at 37 °C. Good signal dispersion in the amide proton and in the aliphatic regions indicated a correctly folded protein. However, after 15 min at 37 °C substantial precipitation with a concomitant deterioration of the NMR spectrum was observed (data not shown).

Clearly, the sample conditions had to be improved, in order to enable long measuring times at 37 °C. Ideally, the sample should show no signs of deterioration even after several days at 37 °C. To improve sample stability, a solution of 1 mM U-Blc was dialyzed into several different buffers and the stability determined by incubation at a series of temperatures for 2 h. The obtained results are summarized in the following table.

	25 °C	30 °C	37 °C
100 mM Tris pH 8, 1 mM EDTA	+	-	-
100 mM Tris pH 8, 1 mM EDTA, 50 mM NaCl	+	~	-
100 mM Tris pH 8, 1 mM EDTA, 150 mM NaCl	+	+	-

	25 °C	30 °C	37 °C
20 mM HEPES pH 7.5	+	~	-
20 mM HEPES pH 7.5, 50 mM NaCl	+	+	-
20 mM HEPES pH 7.5, 150 mM NaCl	+	+	-

	25 °C	30 °C	37 °C
20 mM PO4 pH 6.5	+	+	~
20 mM PO4 pH 6.5, 50 mM NaCl	+	+	~
20 mM PO4 pH 6.5, 150 mM NaCl	+	+	+

	25 °C	30 °C	37 °C
20 mM MES pH 5.5	+	+	+
20 mM MES pH 5.5, 50 mM NaCl	+	+	+
20 mM MES pH 5.5, 150 mM NaCl	+	+	+

	25 °C	30 °C	37 °C
20 mM acetate pH 4.5	+	+	+
20 mM acetate pH 4.5, 50 mM NaCl	+	+	+
20 mM acetate pH 4.5, 150 mM NaCl	+	+	-

	25 °C	30 °C	37 °C
20 mM glycine pH 3.5	+	+	+
20 mM glycine pH 3.5, 50 mM NaCl	+	+	+
20 mM glycine pH 3.5, 150 mM NaCl	+	+	-

Table 2: *Stability of 1 mM Blc in various buffers at different temperatures. Incubation times were 2 h at 25 and 30 °C, and 10 h at 37 °C. Stability was assessed in terms of blurring/precipitation of the sample visible by eye. A "+" sign indicates that the sample was stable, a "~" indicates slight blurring, while a "-" indicates strong precipitation.*

The incubation time at 37 °C was prolonged after the initial 2 h until 10 h. It is evident from the above table that the stability is significantly improved in the more acidic buffers as compared to neutral or basic conditions. The addition of NaCl to the sample showed a weak stabilizing effect at neutral pH and a weak destabilizing effect at acidic pH. We chose the acetate buffer at pH 4.5 to be our sample buffer for any following experiments, if not stated otherwise.

Being able to produce Blc in a reasonable amount and having found the conditions ensuring optimum sample stability, we expressed Blc in ^{15}N-labeled form. Figure 13a illustrates the purification procedure by means of 15% SGS-PAGE. The gel shows that ^{15}N-Blc is > 95% pure. It can also be seen, that the spheroblast pellet still contains a lot of the target protein, i.e. that the osmotic shock lysis is only ~50% effective. An additional faint band with a molecular weight slightly higher than the target protein could not be assigned unambiguously, but probably results from periplasmic Blc, which has its periplasmic signal sequence still attached[64]. The [^{15}N,^{1}H]-HSQC spectrum of ^{15}N-Blc shows good signal dispersion and between 140 and 150 out of an expected 156 peaks (167 residues minus 10 proline residues and the N-terminal residue) (figure 13b).

In order to be able to select the suitable spots in the scaffold protein for grafting of loops and N-terminal segments, the complete resonance assignment of the scaffold molecule would be beneficiary. We therefore, decided to express the Blc protein in a ^{13}C/^{15}N double-labeled form to enable a sequence specific resonance assignment of the peptide backbone by triple-resonance experiments. Standard CBCA(CO)NH[65] and HNCA[66] triple-resonance experiments were recorded as basis for the assignment procedure (figure 13c).

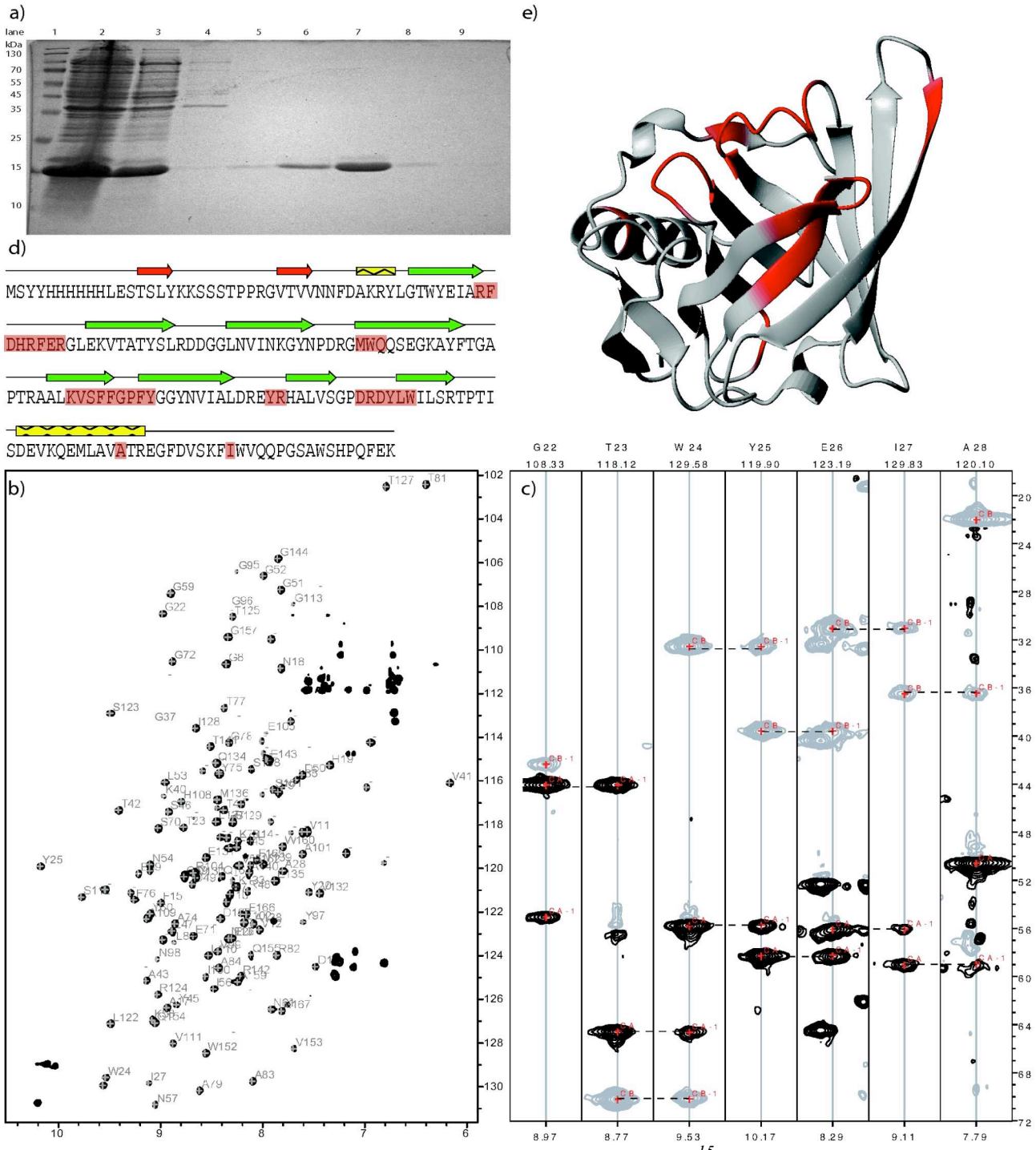


Figure 13: *a) 15% SDS-PAGE of the purification process of ^{15}N-Blc. The lanes contain the following products: 1. pre-stained protein ladder, 2. spheroblast pellet after osmotic lysis, 3. supernatant after osmotic lysis, i.e. crude sample, 4. flow-through of StrepTactin column loading, 5. to 9. eluate fractions 1to 5 b) [^{15}N,^{1}H]-HSQC spectrum with assigned peaks of a ~1 mM ^{15}N-Blc sample in 20 mM d_4-acetate pH 4.5, 10% D_2O, 0.5 mM TMSP recorded at 37 °C. c) Strips from G22 to A28 from the CBCA(CO)NH spectrum of ^{13}C,^{15}N-labeled Blc (same sample conditions as for the [^{15}N,^{1}H]-HSQC) exemplifying the assignment strategy. d) Secondary structural elements observed in the crystal structure of Blc mapped onto the amino acid sequence. Red arrows indicate the strands of the short antiparallel β-sheet lying outside the β-barrel core region. The small yellow rectangle represents the short 3_{10}-helix immediately preceding the β-barrel. β-barrel strands are highlighted with green arrows and the C-terminal α-helix by the large yellow rectangle. Unassigned residues are shaded in red. e) Unassigned regions are colored red in the ribbon representation of the crystal structure of Blc.*

About 80% of the backbone resonances of Blc have been assigned and a table with all assigned chemical shifts can be found in Appendix E. It is interesting to note, that the unassigned residues almost exclusively are located in the odd numbered loops of Blc or in β-strands in the immediate vicinity to these loops. These residues could not be assigned due to very weak signals in the triple-resonance spectra. This is a behavior is often observed in presence of exchange processes, which take place on the microsecond to millisecond timescale.

Chemical exchange processes are temperature dependent and their effect on the NMR resonances additionally depends on the magnetic field strength. Recording spectra at different field strengths and varying temperatures can thus provide clues about the presence of such exchange phenomena. However, [^{15}N,^{1}H]-HSQC spectra recorded from 4 °C to 47 °C at 500 MHz and 700 MHz didn't reveal any such effects (data not shown). Unfortunately higher magnetic field strengths were not accessible for us.

2.3 Design aspects for the soluble scaffold

The N-terminal region of our Blc construct differed from the one for which the crystal structure has been published[10] (figure 4). While our construct – after cleavage of the periplasmic export signal – matches the wildtype Blc almost perfectly, the sequence in the crystal structure carries an additional 19 N-terminal residues, which are a remainder of the chosen cloning strategy. In order to being able to start our grafting trials with a scaffold matching as closely as possible the sequence, for which the threedimensional structure has been published, we decided to appended the amino acid sequence LESTSLYKK to the N-terminus of our Blc construct, as encoded on the plasmid pBlc3 (see also figure 4 above). This construct then served as the starting point for our N-terminal grafting studies. The decision to only append 9 instead of the full 19 residues was supported by the fact that the first 10 residues in the crystal structure are not visible in the crystal structure and are hence supposed to be unstructured and not important for the overall structure of the protein.

2.4 Grafting studies with N-termini of the Y-receptors

Based on the initial success in expression and assignment of both Blc and NY2R, we decided to proceed with our project by grafting the NY2R fragment N-terminally onto the Blc scaffold. Two

attachment points were selected taking into account the lipocalin general architecture (figures 3 and 4). Briefly, the two designed scaffold constructs consisted of (i) exclusively the β-barrel core and the C-terminal α-helix (construct Blc-N2L40 [NY2R grafted before L40 of Blc; Blc numbering refers to the construct used for the crystallographic studies, as deposited under 1QWD.pdb]) and (ii) a longer scaffold containing also the antiparallel β-sheet and the 3_{10}-helix (construct Blc-N2L11).

Blc-N2L40 and Blc-N2L11 could be expressed with yields of 4 and 7 mg/l, respectively. However, during dialysis of Blc-N2L40 a substantial amount of precipitate was observed. In the subsequent 1D ^{1}H-NMR and [^{15}N,^{1}H]-HSQC spectra a low signal intensity in combination with a limited signal dispersion was observed, indicating the presence of denatured protein. Blc-N2L11 however, behaved similar during the purification process than Blc, and NMR samples of ~1 mM concentration showing good signal dispersion could be obtained. Based on the very different outcome from the two grafting attempts we created a series of 6 more attachment point constructs spanning the region between L11 and L40. The additionally chosen sites were T14, S20, T23, P25, N32 and F34. The location of all 8 tested attachment points in the context of the secondary structural elements of the Blc N-terminus are shown in figure 14. They were all evaluated regarding the quality of the resulting NMR spectra (figure 14). The constructs L11 and T14 included all secondary structural elements visible in the crystal structure of Blc and gave rise to satisfactory NMR spectra. Constructs S20, T23 and P25 had the attachment point of NY2R in between the two β-strands forming the short antiparallel β-sheet and thus didn't allow the formation of this secondary structural element. All three constructs showed well-dispersed spectra. It can therefore be concluded that the formation of the short antiparallel β-sheet is not necessary for the formation of the β-barrel, and that this sheet is probably a pure artifact caused by the chosen cloning strategy for the X-ray crystallographic study. Constructs N32 and F34 had the NY2R sequence immediately before the 3_{10}-helix and gave spectra of very poor quality, indicating an interference of the NY2R with the correct formation of the β-barrel. Even worse spectra were observed for the L40 construct, in which the NY2R was cloned immediately before the first β-strand of the barrel. It can thus be concluded that the presence of the short 3_{10}-helix plus a spacer of a few residues is necessary for the formation of the β-barrel core structure of Blc.

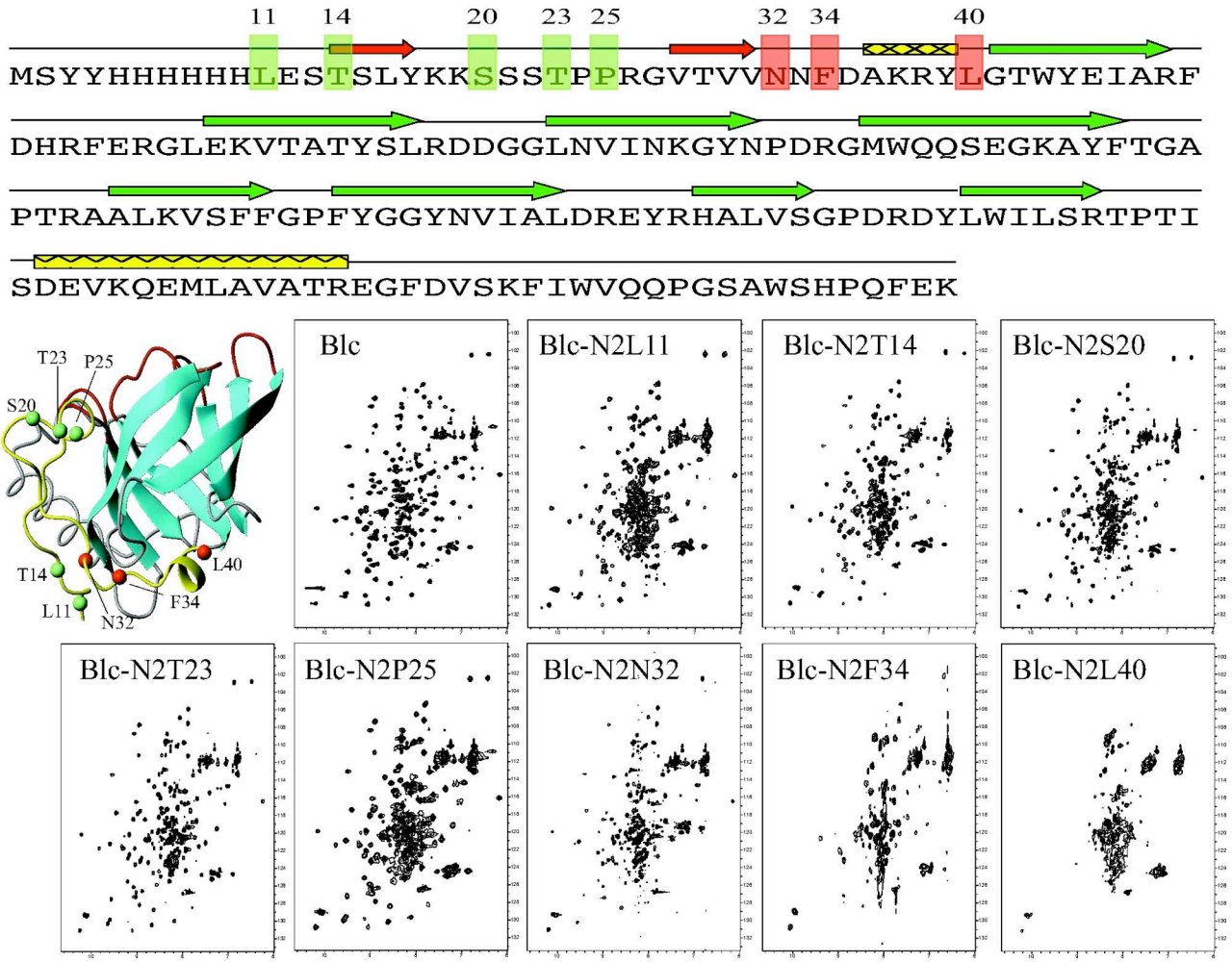


Figure 14: *Top: Secondary structural elements observed in the crystal structure of Blc mapped onto the amino acid sequence. Red arrows indicate the strands of the short antiparallel β-sheet lying outside the β-barrel core region. The small yellow rectangle represents the short 3_{10}-helix immediately preceding the β-barrel. β-barrel strands are highlighted with green arrows and the C-terminal α-helix with a large yellow rectangle. Attachment points of the NY2R are marked with green (tolerant) and red (non-tolerant) shaded areas along with the residue number according to the crystal structure of Blc. Bottom: Crystal structure of Blc with the N-terminus preceding the β-barrel colored in yellow. Attachment points of the NY2R are marked with green (tolerant) and red (non-tolerant) spheres along with the residue number. [^{15}N,^{1}H]-HSQC spectra of Blc as a reference and all described constructs are shown to illustrate, whether an attachment point tolerates the grafting of NY2R or not. All spectra are from ~0.5 mM samples in 20 mM d_4-acetate, pH 4 and were recorded at 27 °C. The spectra range from 99 to 133 ppm and from 6.0 to 10.6 ppm in the ^{15}N and ^{1}H dimensions, respectively.*

In Chapter 1 of this thesis I have described in detail the characterization of all four Y-receptor N-terminal loops. Especially the N-terminus of the Y4-receptor (NY4R) has proofed difficult to produce (Zou, ChemBioChem; 9: 2276-2284, 2008), due to fragmentation of the expressed fusion constructs. It was therefore interesting to find out, if Blc could serve as a scaffold for NY4R and retain stability in solution. NY4R was fused to Blc at the most promising site (S20) identified in the attachment site screening with NY2R. The resulting Blc-N4S20 construct, however, showed fragmentation already

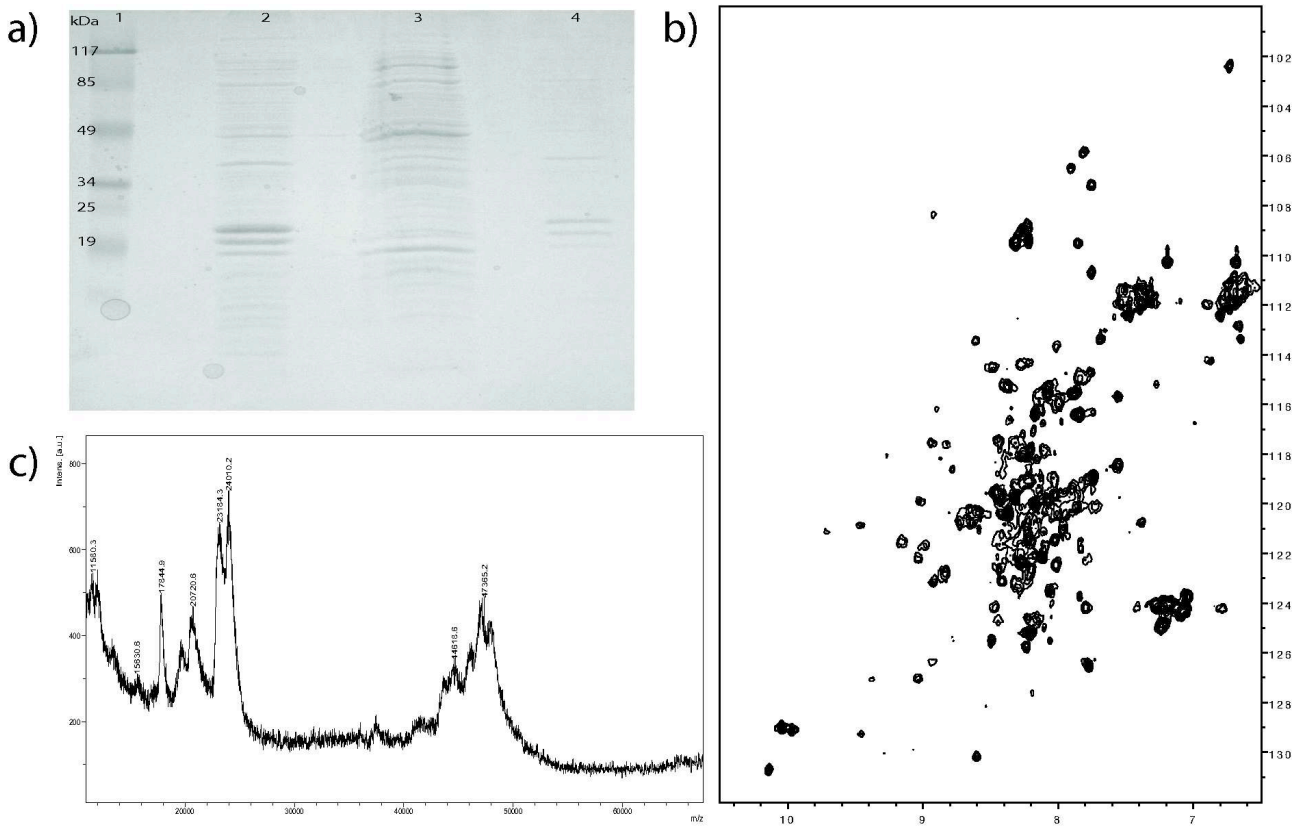


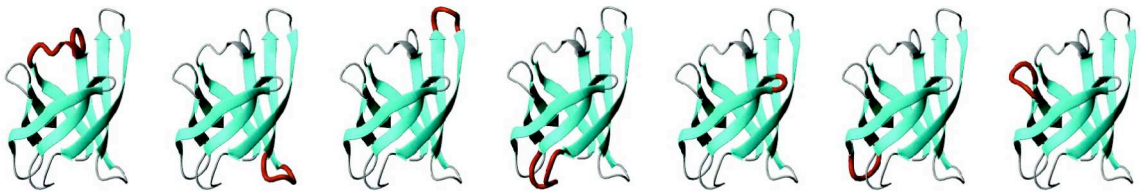
Figure 15: *a) SDS-PAGE of Blc-N4S20 purification. Lane 1 shows a molecular size marker. Lane 2 shows the cells overexpressing Blc-N4S20 after harvesting. Lane 3 is the spheroblast pellet after osmotic lysis of the cells and in lane 4 the final protein preparation after affinity chromatography is shown. The fragmented pattern is retained throughout the expression and purification procedure. b) [^{15}N,^{1}H]-HSQC of 0.5 mM Blc-N4S20 in 20 mM d_4-acetate pH 4.5 recorded at 27 °C. The only strong signals are in the center of the spectrum, which corresponds to the random coil region. c) MALDI-TOF MS of the NMR sample indicates the presence of a broad range of molecular weight species besides the one expected at 23.8 kDa for monomeric Blc-N4S20.*

while in its host cell. The fusion construct could be purified, but NMR spectra were of low quality and mass spectrometry proofed heterogeneity due to fragmentation and aggregation (figure 15).

2.5 Loop grafting studies

In the studies, that used lipocalins as a scaffold, mutations were introduced at the more open side of the β-barrel covering the odd numbered loops 1, 3, 5 and 7. According to the crystal structure of Blc, the residues S20, T23 and P25 are in proximity to these loops. These sites thus represent the attachment points of choice, if the odd numbered loops of Blc should be replaced with Y-receptor extracellular loops. In case one would like to target the even numbered loops 2, 4 and 6 residues L11 and T14 seem to be good candidates for allowing proximity of the grafted N-terminus to the loops.

Residue $D^{6.59}$ of the Y1-receptor was identified as one of the most important for high-affinity ligand binding[67]. It is located in the third extracellular loop of the Y1-receptor. To assess the tolerance of each Blc-loop for being replaced with a Y-receptor loop we generated 7 mutants. In each one native Blc loop was replaced with the extracellular loop 3 of the Y1-receptor (Cys294 was replaced with a serine residue in these constructs). The Y1e3-loop has a length of 14 residues. The length of the Blc loops ranges from 2 to 9 residues. A comparison of all Blc loops with the Y1e3-loop is shown in figure 16. The changes are substantial. However, in previous studies it was found that lipocalins can tolerate large changes to their loop sequences [9].



	d1Y1e3	d2Y1e3	d3Y1e3	d4Y1e3	d5Y1e3	d6Y1e3	d7Y1e3
ΔN	+5	+9	+10	+7	+12	+9	+8
ΔpI	-0.78	+1.76	-	-3.44	-	-0.1	+1.76
Δ-	-1	-1	0	+1	+1	-1	-1
Δ+	-2	-1	-1	-1	0	-2	-1

Figure 16: *Comparison of the Blc loop sequences with the Y1e3-loop sequence. The respective Blc loop is colored red in the ribbon representation of the β-barrel core of Blc (1QWD.pdb). In the table dx denotes the loop being replaced with x representing the loop number. ΔN corresponds to the change in loop length, ΔpI to the change in isoelectric point of the loop sequences and Δ+/- to the change of positively and negatively charged amino acids in the loop.*

Because it is very easy to screen for correctly processed periplasmic proteins, all these mutations have initially been applied to the periplasmic Blc construct. Osmotic shock treatment and analysis of the supernatant didn't show any overexpressed protein. Except for the Blc-d5Y1e3 all the overexpressed protein remained in the spheroblast pellet, which indicated a cytoplasmic location of our constructs. Subsequent lysis of the spheroblasts by sonication also didn't yield any target protein in the supernatant, indicating that our constructs in fact accumulated in the form of insoluble inclusion bodies in cytoplasm (data not shown). Tests with the cytoplasmic expression system led to the same results (data not shown). The inclusion bodies from one construct (Blc-d1Y1e3) were purified and analyzed by SDS-PAGE and MS. Both indicated the presence of two species of slightly different molecular weight. Interestingly the two species corresponded to the Blc-d1Y1e3 mutant with and without the periplasmic export signal, which indicates aggregation and deposit into inclusion bodies after transport to the

plasmamembrane and interaction with the signal peptidase.

As mentioned, loop 5 of Blc seemed to be the only one, which is able to tolerate exchange with the foreign Y1e3 loop. It could be shown that this tolerance is generic by also replacing the loop with the extracellular loops 1 and 2 of the Y1-receptor (Y1e1 and Y1e2, with Cys197 mutated to serine). However, precipitate formation during the purification of these three constructs indicated a decreased stability compared to wt-Blc. This was corroborated by NMR spectra, which showed very weak signals and rapid sample deterioration (data not shown). Other loop positions were almost completely intolerant to changes in their sequences. This was especially striking, when we tried to exchange those loops with sections from the Y1-receptor extracellular loops, which were of equal length as the Blc loop sequences or equal-length sequences composed solely of glycines and serines.

Several strategies were tested, to obtain the Blc mutants described above in stable folded forms. Growing the cultures under more "gentle" conditions didn't give rise to any folded target protein. Likewise he inclusion of osmolytes, such as betaine and benzylic alcohol, in the growth medium didn't show any beneficial effects. Several different refolding strategies, including rapid dilution, slow dilution by dialysis and on-column methods each carried out under various different conditions, didn't yield any stably folded protein. We therefore concluded that the unstructured nature of our constructs was not due to incorrect processing in the cell, but rather due to an inherent incapacity of the given amino acid sequences to adopt a stable fold.

3. Discussion

Structural information for G-protein coupled receptors (GPCRs) is hard to obtain and therefore still scarce. Difficulties associated with their study partly are inherent to all membrane proteins and partly are particular to the GPCRs. New methods besides the established X-ray crystallographic methods might thus yield valuable insight into the structure and function of these receptors of tremendous biological importance. Two such methods, namely the characterization of fragments of the receptor and the grafting of functionally important epitopes onto a soluble scaffold molecule, have been applied to the neuropeptide Y receptors (Y-receptors), a subfamily of the GPCRs. We were able to produce the N-terminus of the Y2-receptor (NY2R) both in unlabeled and in ^{15}N-labeled form as a soluble fusion to ubiquitin. After specific cleavage with the yeast ubiquitin hydrolase NY2R could be characterized by NMR spectroscopy. The 50 residue peptide didn't show any defined tertiary or secondary structural

elements, both in aqueous buffer and in DPC micelles. This is in contrast to other Y-receptor N-termini[49,51]. Chemical shift mapping data indicate a weak interaction of NY2R with the natural ligands of the Y-receptors (NPY, PYY and PP). Such weak interactions with their ligands could be observed for all four Y-receptor N-termini. Whereas with two such N-termini stability problems have been encountered during expression[49,51], this was not the case for NY2R. The NY2R seemed therefore a good candidate for grafting onto a soluble scaffold protein, such as the bacterial lipocalin (Blc). Blc has been cloned and purified in other laboratories and its structure has been solved by X-ray crystallography [10]. However, no NMR data have been published on Blc. Our efforts to produce Blc in ^{15}N- and ^{15}N,^{13}C-labeled forms have enabled the acquisition of high-quality heteronuclear NMR spectra. In the [^{15}N,^{1}H]-NMR spectrum of Blc out of the 156 expected peaks (167 residues minus N-terminal residue and 10 prolines), 140-150 peaks can be observed. Out of these ~125 backbone amide correlations could be assigned. The remaining 15-25 peaks showed very weak signals in the triple-resonance spectra recorded for assignment purposes. This is often observed, if resonances are broadened by exchange processes on the millisecond to microsecond time scale. As the resonances that could not be assigned, are almost exclusively located on one side of Blc's β-barrel core, it can be assumed that these residues show some flexibility and are neither in highly restricted conformations, nor are they highly dynamic.

In the crystal structure of Blc [10] a short antiparallel β-sheet and a short 3_{10}-helix are found in the N-terminus domain preceding the β-barrel core. The NY2R sequence was appended to the N-terminus of the β-barrel core at various attachment points distributed throughout the Blc N-terminus. It was found, that some attachment points were well tolerated by the scaffold, whereas other interfered with proper folding of the β-barrel core. Generally it can be said, that the further away from the β-barrel core, the better the attachment of the foreign NY2R sequence was tolerated. Attachment immediately prior to the β-barrel core, eliminating both the short β-sheet and the 3_{10}-helix, was not tolerated, suggesting the importance of either the helix, the sheet or both for the overall structure of the protein. Disrupting the β-sheet, but leaving the 3_{10}-helix in place, was tolerated by the scaffold, suggesting the insignificance of the β-sheet for the overall structure of the protein. This is in accordance with the fact that the 3_{10}-helix is a conserved structural feature found in many lipocalins[6,17] whereas the short antiparallel β-sheet is not. As the N-terminal strand of the β-sheet is made up of residues artificially introduced into the construct through the chosen cloning strategy [10], its dispensability is also not unanticipated.

Whereas the attachment of the NY2R sequence to the N-terminus of Blc could be achieved without impairing the formation of the β-barrel core, the loops connecting the strands of the barrel were much less tolerant to being replaced with foreign sequences. This was surprising considering the fact that the

loops in the lipocalins show little sequence conservation among different members of the lipocalin family[9]. Furthermore these loops have been subject to extensive mutagenesis in the production of so-called anticalins[11,68]. However, the anticalins were generated using phage display techniques, in which randomized proteins in the loops were generated. Binders for a certain target molecule were then enriched in several panning rounds, as is common practice in phage display techniques. This setting most likely not only allows enriching a population of specific binders for the target molecule. As such binding is probably only possible in cases, where the canonical structure of the protein core is retained, the technique at the same time also selects for mutations, which don't interfere with this fold. Our approach, namely the introduction of a particular amino acid sequence, was fundamentally different from the random mutagenesis approach used in the phage display strategy. We conclude that even though the lipocalins can be used as molecular scaffolds, their usefulness in this respect is likely linked to phage display techniques.

The failure to generate a soluble minireceptor prompted us for trying out a membrane protein, which had been used successfully as a scaffold previously. The results resulting from this project will be summarized in Chapter 4 of this thesis.

4. Materials and Methods

4.1 Cloning and expression of NY2R

See Materials and Methods in chapter II.

4.2 Expression of neurohormones

See Materials and Methods in chapter II.

4.3 Plasmids for the overexpression of Blc

Plasmids for the periplasmic (pBlc3) as well as the cytoplasmic (pBlc5) expression of Blc were

available to us. They were derived from the pASK-IBA2 plasmid. Maps of the expressed region of the two plasmids can be found in Appendix F.

4.4 Conditions for the growth of cells expressing Blc and its derivatives

All cultures were grown at 22 °C and 180 rpm and were induced with anhydrotetracyclin to a final concentration of 0.2 µg/ml from a 2 mg/ml anhydrotetracyclin stock in N,N-dimethylformamide once the OD_{550} had reached 0.5. Cells were harvested once the cultures reached stationary phase by centrifugation at 4 °C and stored at -20 °C until further processing.

4.5 Expression of unlabeled Blc

To assess efficiency of BL21 DE3 and C41 DE3 for the cytoplasmic and periplasmic expression of Blc, 50 ml of LB containing 100 µg/ml ampicillin were inoculated in a 1:100 dilution with an overnight LB-culture of BL21 DE3 pBlc3, BL21 DE3 pBlc5, C41 DE3 pBlc3 and C41 DE3 pBlc5. The cultures were incubated in 300 ml Erlenmeyer flasks. After 24 h the cultures had reached the stationary phase at an OD_{550} of 2.3 and were harvested. For each one of the cultures a yield of cell wet biomass of approximately 6 g/l was obtained.

4.6 Expression of ^{15}N-labeled Blc

1 liter of minimal medium (composition see Appendix A) was inoculated in a 1:200 dilution with a saturated LB overnight pre-culture and distributed to three 2 l Erlenmeyer flasks. After 11 h the cultures were induced. After 25 h the cultures had reached an OD_{550} of 2.3 and were harvested to yield a wet biomass of 3.2 g from 1 liter of culture.

4.7 Expression of ^{13}C/^{15}N-labeled Blc

10 ml of LB-broth containing 100 µg/ml ampicillin were inoculated with a colony of freshly transformed BL21 DE3 pBlc3 and incubated at 37 °C and 240 rpm until the OD_{550} was ~0.8. This pre-

culture was centrifuged at room temperature for 10 min at ~1000 g. The resulting approx. 50 µl pellet was resuspended in 50 ml $^{13}C/^{15}N$-labeled minimal medium (composition see Appendix B) containing 3 g ^{13}C-Glucose per liter and incubated at 22 °C and 180 rpm until blurring of the culture was visible by eye (OD_{550} ~0.25). This 50 ml pre-culture was used to inoculate the remaining 950 ml of $^{13}C/^{15}N$-labeled minimal medium, which was then distributed to three 2 l Erlenmeyer flasks and incubated at 22 °C and 180 rpm. After 6 h the cultures were induced. After 21 h the cultures reached an OD_{550} of 2.5 and were harvested to yield a wet biomass of 4.6 g from 1 liter of culture.

4.8 Purification of Blc and its derivatives

Cell pellets were thawed for 1 h on ice before 3.3 ml of chilled buffer P (0.5 M sucrose, 100 mM Tris pH 8, 1 mM EDTA) supplemented with 100 µg/ml lysozyme was added per 1 g of wet biomass. The viscous solution was incubated on ice for 30-60 min and centrifuged for 30 min at 38700 g at 4 °C. The supernatant was centrifuged a second time for 30 min at 38700 g at 4 °C to yield a clear, viscous, slightly green colored solution. All subsequent purification steps were carried out at 4 °C. The supernatant was loaded onto a StrepTactin column (1 ml of StrepTactin resin per g of cell pellet), previously equilibrated with 5 column volumes (CV) of buffer W (100 mM Tris pH 8, 1 mM EDTA), by means of a peristaltic pump at a flow-rate of approximately 0.3 ml/min. Bound Blc was washed with 5 CV of buffer W, before it was eluted with 5 CV of buffer E (100 mM Tris pH 8, 1 mM EDTA, 2.5 mM desthiobiotin). Fractions were collected and analyzed for content of the target protein by 15% SDS-PAGE.

The column was regenerated by washing away bound desthiobiotin with 10 CV of buffer R (100 mM Tris pH 8, 1 mM EDTA, 5 mM 2-(4-Hydroxyphenylazo)benzoic acid [HABA]) and equilibrating it with 10 CV of buffer W.

The Blc-containing eluate fractions were pooled and the concentration estimated from the UV-absorption at 280 nm using a theoretical absorption coefficient e_{280} of 40910 $M^{-1}cm^{-1}$. The Blc solution in the elution buffer was transferred into a 3.5 kDa MW cutoff dialysis bag previously soaked in 20 mM acetate buffer pH 4.5 at 4 °C. The solution was dialyzed twice against typically a 100-200 fold excess of 20 mM acetate buffer pH 4.5 at 4 °C during 12 h. The dialyzed protein solution was centrifuged for 10 min at 2500 g at 4 °C to remove any residual precipitate that had formed during the dialysis before being concentrated to a volume suitable for NMR (0.25-0.6 ml) in a Amicon Ultra-4 centricon tube (10 kDa MWCO) at 4 °C. The dialysis buffer was exchanged for the desired NMR

buffer (20 mM ^{2}H-acetate buffer pH 4.5, 10% ^2H_2O, 0.5 mM TSP) in the same centricon tube in three 1:1 dilution steps. The resulting solution was carefully removed from the centricon tube and centrifuged in an Eppendorf tube at 12000 rpm at room temperature for 2 min before being transferred to a clean NMR tube.

4.9 Grafting of Y-receptor N-termini

The plasmids coding for the constructs Blc-N2L11, -N2T14, -N2S20, -N2N32, -N2F34 and -N2L40 carrying the N-terminus of the Y2-receptor at various sites were generated from pBlc3 (see Appendix F for a plasmid map) as the starting plasmid. They were all generated by an overlapping PCR strategy consisting of three rounds for each construct. In summary the Y-receptor N-terminal fragment had to be inserted in between the OmpA periplasmic signal sequence at the N-terminus and the Blc β-barrel core at the C-terminus. This was achieved by generating by PCR three constructs comprising (i) the *Xba*I restriction for cloning site and the OmpA signal sequence, (ii) the Y-receptor N-terminus and (iii) the Blc core plus a *Hind*III restriction site for cloning. All the fragments were designed to show overlap on both sides with the preceding and following fragment as illustrated in the following figure.

A table with the primers used to generate the three respective fragments of all 6 constructs is shown below (all primers are depicted as going from the 5' to the 3' end).

N2L11	**Fragment 1**	**Fragment 2**	**Fragment 3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	ctggaatctactagcctgtacaaaa aaagttctcctacgccgccg
rv	gatcggacccatggcggcctgcg ctacggt	ttttttgtacaggctagtagattccag ctgaacttcgatcagttt	tcacttcacaggtcaagctta
N2T14	**Fragment1**	**Fragment2**	**Fragment3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	actagcctgtacaaaaaaagttctc ctacgccgccg
rv	gatcggacccatggcggcctgcg ctacggt	ttttttgtacaggctagtctgaacttc gatcagttt	tcacttcacaggtcaagctta
N2S20	**Fragment1**	**Fragment2**	**Fragment3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	atcgaagttcagagttctcctacg
rv	gatcggacccatggcggcctgcg ctacggt	cgtaggagaactctgaacttcgat	tcacttcacaggtcaagctta
N2N32	**Fragment1**	**Fragment2**	**Fragment3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	atcgaagttcagaataatttcgac
rv	gatcggacccatggcggcctgcg ctacggt	gtcgaaattattctgaacttcgat	tcacttcacaggtcaagctta

N2F34	**Fragment1**	**Fragment2**	**Fragment3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	atcgaagttcagttcgacgccaac
rv	gatcggacccatggcggcctgcg ctacggt	gttggcgtcgaactgaacttcgat	tcacttcacaggtcaagctta
N2L40	**Fragment1**	**Fragment2**	**Fragment3**
fw	cgacaaaaatctagataacgaggg	gcgcaggccgccatgggtccgat cggcgcg	cttggtacctggtatgagattgcc
rv	gatcggacccatggcggcctgcg ctacggt	ggcaatctcataccaggtaccaag ctgaacttcgatcagttt	tcacttcacaggtcaagctta

The fragments were generated by standard PCR procedures using Vent polymerase PCR products were analyzed and purified by 1.5% agarose gel electrophoresis containing ethidium bromide for DNA staining. A QiaGen gel purification kit was used for all PCR purifications following the manual of the manufacturer.

200 to 500 ng of the final constructs were digested with *Xba*I (5 U) and *Hind*III (10 U) in Tango buffer at 37 °C for 2 h, purified on a 1.5% agarose gel and ligated into the pBlc3 vector, digested with *Xba*I and *Hind*III and purified previously. The ligation reaction was transformed to DH5α cells. Plasmid was extracted from positive clones using the QiaGen miniprep kit following the manufacturers instructions, and the sequence of the insert was verified by dideoxy-sequencing (Sanger, PNAS, 74, 5463, 1977).

The constructs Blc-N2T23, -N2P25 and -N4S20 were constructed with following the QuickChange mutagenesis strategy. For the two constructs carrying the N-terminus of the Y2-receptor (NY2R) pBlc3-N2S20 was used as the starting construct from which 3 and 5 residues between the NY2R sequence and the Blc sequence were deleted to generate pBlc3-N2T23 and pBlc3-N2P25, respectively.

The primers used for the generation of pBlc3-N2T23 were 5'-ctgatcgaagttcagacgccgccgcgtggcg-3'for the forward and 5'-cgccacgcggcggcgtctgaacttcgatcag-3' for the reverse, and for the generation of pBlc3-N2P25 5'-ccaaactgatcgaagttcagccgcgtggcgtgacc-3' for the forward and 5'-ggtcacgccacgcggctgaacttcgatcagtttgg-3' for the reverse primer.

pBlc3-N4S20 was generated from pBlc3 as the starting plasmid. QuickChange primers for this construct were generated from two short primers (forward primer: 5'-gctaccgtagcgcaggccgccATGAACACCTCTCACC-3'; reverse primer: 5'-ccacgcggcggcgtaggagaactCATCACGTCCACGGAATCC-3'; lowercase letters denote stretches from the Blc sequence, uppercase letters stretches from the NY4R) amplifying the N-terminus of the Y4-receptor (NY4R) with overhangs upstream and downstream of the S20 site on both sides. ds-DNA fragments were generated by standard PCR methods using Vent polymerase and the fragments were purified using a QiaGen gel purification kit following the manual of the manufacturer. QuickChange

PCR reactions were carried out following the recommendations from Stratagene (for more detailed description of the protocols followed *vide infra*). Nucleotide sequences of the inserts were verified using dideoxy-sequencing[69].

4.10 Grafting of Y-receptor extracellular loops

All designed constructs were derived from the pBlc3 or pBlc5 plasmid and all mutants were created using the QuickChange methodology as described in the QuickChange site directed-mutagenesis kit manual from Stratagene (catalog no. 200518). Primer design principles suggested in this manual were followed. All components of the kit were bought individually. The dNTP were from Fermentas (cat. no. #R0241) and were stored as aliquots at -20 °C. PfuTurbo polymerase was from Stratagene (cat. no. 600250-52) and DpnI nuclease was from Fermentas (cat. no. #ER1701). All QuickChange reactions were carried out in 10 μl total volume containing 1x PfuTurbo buffer (cat. no. 600153-82), 0.5 U PfuTurbo polymerase, 0.2 mM dNTPs, ~100 ng of plasmid DNA and 0.2 μM of each mutagenic primer. The thermocycling procedure was designed as recommended by Stratagene and can be found in appendix B. 5 U of DpnI were added to the reaction and the mixture was incubated at 37 °C for 3 hours. 1-2 μl of this mixture were transformed to 25-50 μl of self-prepared $RbCl_2$-competent cells by heat shock. Plasmid was extracted from the obtained colonies and sequenced by the dideoxy sequencing method (Sanger, PNAS, 74, 5463, 1977).

All primers were purchased from Microsynth (Balgach, Switzerland). Primers for deletion, single amino acid change and point mutation reactions were purchased as desalted ~36-mers. Primers for the insertion constructs were self-made by PCR using two short (~12 nucleotides) desalted primers flanking the insertion region. PCR products were purified with a Sigma PCR clean-up kit (NA1020-1KT) and used in subsequent QuickChange mutagenesis reaction. Since long primers are expensive and require additional expensive purification, this strategy allowed us substantial cost savings. The inserted Y-receptor gene sequences were optimized for preferred *E. coli* codon usage. A table with all the deletion and insertion primers is shown in Appendix G.

4.11 NMR spectroscopy

4.11.1 NMR spectroscopy - general

All spectra were recorded on a Bruker AV-600 or AV-700 spectrometer, both equipped with cryoprobes.

Proton chemical shifts were calibrated to the water signal and nitrogen shifts were referenced indirectly to liquid NH_3 (Live, 1984). Raw data was processed using the Bruker Topspin software version 2.0 and 2.1 and transferred to XEASY[70] or CARA[71] for further analysis.

All 2D experiments utilized TPPI-States[72,73] for quadrature detection in indirect proton and carbon dimensions, and gradient-selected coherence selection (echo-antiecho) (Kay, JACS, 114, 10663, 1992) in combination with sensitivity enhancement schemes[74,75] in experiments including detection of amide protons.

4.11.2 Assignment of NY2R

See Materials and Methods in chapter II.

4.11.3 Assignment of Blc

All spectra for the assignment of Blc were recorded on the AV-600 spectrometer on a 1.1 mM sample of uniformly ^{15}N- and ^{13}C-labeled Blc in 20 mM d_4-acetate at pH 4.5 containing 10% D_2O and at a temperature of 308 K.

A [^{15}N,^{1}H]-HSQC[76] with 1024(^{1}H)*150(^{15}N) complex data points was recorded. Spectral widths were 16 ppm and 40 ppm for the ^{1}H and ^{15}N dimension, respectively. The carrier positions were placed at 4.75 ppm (^{1}H) and 115.0 ppm (^{15}N).

A CBCA(CO)NH[77,78] with 1024(^{1}H)*32(^{15}N)*75(^{13}C) complex data points was recorded. Spectral widths were 14 ppm, 35 ppm and 75 ppm for the ^{1}H, ^{15}N and ^{13}C dimension, respectively. The carrier positions were placed at 4.71 ppm (^{1}H), 117.4 ppm (^{15}N), 39.0 ppm ($^{13}C_{\alpha/\beta}$) and 54.0 ppm ($^{13}C_{\alpha}$).

A HNCACB [78,79] with 1024(^{1}H)*32(^{15}N)*64(^{13}C) complex data points was recorded. Spectral widths were 14 ppm, 35 ppm and 75 ppm for the ^{1}H, ^{15}N and ^{13}C dimension, respectively. The carrier positions were placed at 4.71 ppm (^{1}H), 117.4 ppm (^{15}N) and 39.0 ppm (^{13}C).

A (H)CC(CO)NH[80-83] with 1024(^{1}H)*20(^{15}N)*64(^{13}C) complex data points was recorded. Spectral widths were 16 ppm, 35 ppm and 75 ppm for the ^{1}H, ^{15}N and ^{13}C dimension, respectively. The carrier

positions were placed at 4.71 ppm (^{1}H), 117.4 ppm (^{15}N), 39.0 ppm (^{13}C) and 54.0 ppm ($^{13}C_{\alpha}$).
A (H)N(CA)NH[84] with 1024(^{1}H)*24(^{15}N)*55(^{15}N) complex data points was recorded. Spectral widths were 16 ppm, 36 ppm for the ^{1}H and the two ^{15}N dimensions, respectively. The carrier positions were placed at 4.76 ppm (^{1}H), 117.5 ppm (^{15}N).

4.11.4 NMR spectroscopy with Blc derivatives carrying a Y-receptor N-terminus or loop
All protein samples were studied by measuring [^{15}N,^{1}H]-HSQCs with 1024(^{1}H)*150(^{15}N) complex data points was recorded. Spectral widths were 16 ppm and 40 ppm for the ^{1}H and ^{15}N dimension, respectively. The carrier positions were placed at 4.75 ppm (^{1}H) and 115.0 ppm (^{15}N).

5. References

1. Ma, P. & Zemmel, R. Value of novelty? *Nat Rev Drug Discov* **1**, 571-572 (2002).
2. Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr Rev* **21**, 90-113 (2000).
3. Sollod, B. L. et al. Were arachnids the first to use combinatorial peptide libraries? *Peptides* **26**, 131-139 (2005).
4. Pervaiz, S. & Brew, K. Homology of β-lactoglobulin, serum retinol-binding protein, and protein HC. *Science* **228**, 335-337 (1985).
5. Akerstrom, B., Flower, D. R. & Salier, J. P. Lipocalins: unity in diversity. *Biochim Biophys Acta* **1482**, 1-8 (2000).
6. Flower, D. R. The lipocalin protein family: structure and function. *Biochem J* **318**, 1-14 (1996).
7. Flower, D. R., North, A. C. & Sansom, C. E. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta* **1482**, 9-24 (2000).
8. Skerra, A. Engineered protein scaffolds for molecular recognition. *J Mol Recognit* **13**, 167-187 (2000).
9. Skerra, A. Lipocalins as a scaffold. *Biochim Biophys Acta* **1482**, 337-350 (2000).
10. Campanacci, V. et al. The crystal structure of the Escherichia coli lipocalin Blc suggests a possible role in phospholipid binding. *FEBS Lett* **562**, 183-188 (2004).
11. Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci U S A* **96**, 1898-1903 (1999).
12. Korndorfer, I. P., Schlehuber, S. & Skerra, A. Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin. *J Mol Biol* **330**, 385-396 (2003).
13. Schlehuber, S. & Skerra, A. Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant -- a so-called 'anticalin' -- using a molecular random approach. *Biophys Chem* **96**, 213-228 (2002).
14. Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine

antibody fragment in Escherichia coli. *Gene* **151**, 131-135 (1994).
15. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356-363 (2009).
16. Boivin, S. et al. Characterization of Urotensin-II Receptor Structural Domains Involved in the Recognition of U-II, URP, and Urantide. *Biochemistry* **45**, 5993-6002 (2006).
17. Bishop, R. E. The bacterial lipocalins. *Biochim Biophys Acta* **1482**, 73-83 (2000).
18. Bader, R., Lerch, M. & Zerbe, O. BioNMR in Drug Research. 95-120 (2002).
19. Lerch, M. et al. Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J. Mol. Biol.* **322**, 1117-1133 (2002).
20. Lerch, M. et al. Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding. *Biochemistry* **44**, 9255 - 9264 (2005).
21. Lerch, M., Mayrhofer, M. & Zerbe, O. Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J. Mol. Biol.* **339**, 1153-1168 (2004).
22. Beck-Sickinger, A. G. & Jung, G. Structure-activity relationships of neuropeptide Y analogues with respect to Y1 and Y2 receptors. *Biopolymers* **37**, 123-142 (1995).
23. Beck-Sickinger, A. G. et al. Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur. J. Biochem.* **225**, 947-958 (1994).
24. Sautel, M. et al. Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. *Mol. Pharmacol.* **50**, 285-292 (1996).
25. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J. Biol. Chem.* **269**, 2863-2869 (1994).
26. Herzog, H. et al. Cloned human neuropeptide Y receptor couples to two different second messenger systems. *Proc Natl Acad Sci U S A* **89**, 5794-5798 (1992).
27. Krause, J., Eva, C., Seeburg, P. H. & Sprengel, R. Neuropeptide Y1 subtype pharmacology of a recombinantly expressed neuropeptide receptor. *Mol Pharmacol* **41**, 817-821 (1992).
28. Larhammar, D. et al. Cloning and functional expression of a human neuropeptide Y/peptide YY receptor of the Y1 type. *J Biol Chem* **267**, 10935-10938 (1992).
29. Rose, P. M. et al. Cloning and functional expression of a cDNA encoding a human type 2 neuropeptide Y receptor. *J Biol Chem* **270**, 22661-22664 (1995).
30. Bard, J. A., Walker, M. W., Branchek, T. A. & Weinshank, R. L. Cloning and functional expression of a human Y4 subtype receptor for pancreatic polypeptide, neuropeptide Y, and peptide YY. *J Biol Chem* **270**, 26762-26765 (1995).
31. Lundell, I. et al. Cloning of a human receptor of the NPY receptor family with high affinity for pancreatic polypeptide and peptide YY. *J Biol Chem* **270**, 29123-29128 (1995).
32. Gerald, C. et al. A receptor subtype involved in neuropeptide-Y-induced food intake. *Nature* **382**, 168-171 (1996).
33. Blomqvist, A. G. & Herzog, H. Y-receptor subtypes--how many more? *Trends Neurosci* **20**, 294-298 (1997).
34. Larhammar, D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **65**, 165-174 (1996).
35. Robin-Jagerschmidt, C. et al. The ligand binding site of NPY at the rat Y1 receptor investigated by site-directed mutagenesis and molecular modeling. *Mol Cell Endocrinol* **139**, 187-198 (1998).
36. Wieland, H. A., Eckard, C. P., Doods, H. N. & Beck-Sickinger, A. G. Probing of the neuropeptide Y-Y1-receptors interaction with anti-receptor antibodies. *Eur J Biochem* **255**, 595-603 (1998).
37. Clark, J. T., Kalra, P. S., Crowley, W. R. & Kalra, S. P. Neuropeptide Y and human pancreatic

polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427-429 (1984).
38. Nakajima, M. et al. Effects of pancreatic polypeptide family peptides on feeding and learning behavior in mice. *J Pharmacol Exp Ther* **268**, 1010-1014 (1994).
39. Larhammar, D. Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11 (1996).
40. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A* **78**, 4175-4179 (1981).
41. Li, X. A., Sutcliffe, M. J., Schwartz, T. W. & Dobson, C. M. Sequence-specific ^{1}H NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **31**, 1245-1253 (1992).
42. Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R. & Zerbe, O. Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution:‚Äâ Determinants of Stability of a Helical Hairpin. *Journal of the American Chemical Society* **129**, 8811-8817 (2007).
43. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R. J. Solution structure of monomeric peptide YY supports the functional significance of the PP-fold. *Biochemistry* **39**, 9935-9942 (2000).
44. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. Solution structure of human neuropeptide Y. *J Biomol NMR* **8**, 379-390 (1996).
45. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection. *J. Mol. Biol.* **305**, 307-392 (2001).
46. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3**, 842-848 (1996).
47. Kuliopulos, A. & Walsh, C. Production, Purification, and Cleavage of Tandem Repeats of Recombinant Peptides. *J. Am. Chem. Soc,* **116**, 4599-4607 (1994).
48. Bornstein, P. & Balian, G. Cleavage at Asn-Gly bonds with hydroxylamine. *Methods Enzymol* **47**, 132-145 (1977).
49. Zou, C., Kumaran, S., Walser, R. & Zerbe, O. Properties of the N-terminal domains from Y receptors probed by NMR spectroscopy. *J Pept Sci* **15**, 184-191 (2009).
50. Wüthrich, K. NMR of Proteins and Nucleic Acids. (1986).
51. Zou, C., Kumaran, S., Markovic, S., Walser, R. & Zerbe, O. Studies of the structure of the N-terminal domain from the Y4 receptor - a G protein-coupled receptor - and its interaction with hormones from the NPY family. *Chembiochem* **9**, 2276-2284 (2008).
52. Movva, N. R., Nakamura, K. & Inouye, M. Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of Escherichia coli. *J Biol Chem* **255**, 27-29 (1980).
53. Muller, H. N. & Skerra, A. Functional expression of the uncomplexed serum retinol-binding protein in Escherichia coli. Ligand binding and reversible unfolding characteristics. *J Mol Biol* **230**, 725-732 (1993).
54. Wulfing, C. & Pluckthun, A. Protein folding in the periplasm of Escherichia coli. *Mol Microbiol* **12**, 685-692 (1994).
55. Skerra, A. & Pluckthun, A. Secretion and in vivo folding of the Fab fragment of the antibody McPC603 in Escherichia coli: influence of disulphides and cis-prolines. *Protein Eng* **4**, 971-979 (1991).
56. Sutton, M. R. et al. Amino acid sequence of Escherichia coli biotin carboxyl carrier protein (9100). *J Biol Chem* **252**, 3934-3940 (1977).
57. Korpela, M. T., Kurittu, J. S., Karvinen, J. T. & Karp, M. T. A recombinant Escherichia coli sensor strain for the detection of tetracyclines. *Anal Chem* **70**, 4457-4462 (1998).
58. Loferer, H., Hammar, M. & Normark, S. Availability of the fibre subunit CsgA and the nucleator

protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. *Mol Microbiol* **26**, 11-23 (1997).
59. Schmidt, T. G. & Skerra, A. The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Eng* **6**, 109-122 (1993).
60. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc Natl Acad Sci U S A* **98**, 3750-3755 (2001).
61. Schmidt, T. G., Koepke, J., Frank, R. & Skerra, A. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol* **255**, 753-766 (1996).
62. Voss, S. & Skerra, A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng* **10**, 975-982 (1997).
63. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-298 (1996).
64. Tuteja, R. Type I signal peptidase: an overview. *Arch Biochem Biophys* **441**, 107-111 (2005).
65. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc,* **114**, 6291-6293 (1992).
66. Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-4667 (1990).
67. Merten, N. et al. Receptor subtype-specific docking of Asp6.59 with C-terminal arginine residues in Y receptor ligands. *J Biol Chem* **282**, 7543-7551 (2007).
68. Vogt, M. & Skerra, A. Construction of an artificial receptor protein ("anticalin") based on the human apolipoprotein D. *Chembiochem* **5**, 191-199 (2004).
69. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
70. Bartels, C., Xia, T. H., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **5**, 1-10 (1995).
71. Keller, R. The Computer Aided Resonance Assignment. (2004).
72. Bodenhausen, G., Vold, R. L. & Vold, R. R. Multiple quantum spin-echo spectroscopy. *J Magn Reson* **37**, 93-106 (1980).
73. Marion, D. & Wuthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun* **113**, 967-974 (1983).
74. Kay, L. E., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc,* **114**, 10663-10665 (1992).
75. Palmer, A. G. I. I. I., Cavanagh, J., Wright, P. E. & Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson* **93**, 151-170 (1991).
76. Bodenhausen, G. Heteronuclear J spectroscopy. *Chem. Phys. Lett.* **39**, 175-179 (1980).
77. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J Biomol NMR* **3**, 185-204 (1993).
78. Muhandiram, D. R. & Kay, L. E. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J Magn Reson B* **103**, 203-216 (1994).
79. Wittekind, M. & Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the α- and β-Carbon Resonances in Proteins. *J*

Magn Reson **101**, 201-205 (1993).
80. Clowes, R. T., Boucher, W., Hardman, C. H., Domaille, P. J. & Laue, E. D. A 4D HCC(CO)NNH experiment for the correlation of aliphatic side-chain and backbone resonances in $^{13}C/^{15}N$-labelled proteins. *J Biomol NMR* **3**, 349-354 (1993).
81. Grzesiek, S., Anglister, J. & Bax, A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in $^{13}C/^{15}N$-Enriched Proteins by Isotropic Mixing of ^{13}C Magnetization. *J Magn Reson* **101**, 114-119 (1993).
82. Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J Biomol NMR* **3**, 225-231 (1993).
83. Montelione, G. T., Lyons, B. A., Emerson, S. D. & Tashiro, M. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc,* **114**, 10974-10975 (1992).
84. Weisemann, R., Rüterjans, H. & Bermel, W. 3D Triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins. *J Biomol NMR* **3**, 113-120 (1993).

6. Appendix

A - 1 liter ^{15}N-labeled minimal medium:

4 g KH_2PO_4

4 g K_2HPO_4

3.5 g $Na_2HPO_4 * 2\ H_2O$

1 g NaCl

1 g $^{15}NH_4Cl$ (or $^{14}NH_4Cl$ if no labeling was required)

fill up to 960 ml with de-ionized water and autoclave

To these 960 ml the following were added:

10 ml $MgSO_4$ 1 M (autoclaved)

25 ml Glucose 20% (sterile filtered)

2 ml trace metal stock solution (autoclaved) (for composition see Appendix C)

1 ml thiamineHCl 150 mM (sterile filtered)

1 ml kanamycin 50 mg/ml or ampicillin 100 mg/ml

B - 1 liter $^{13}C/^{15}N$-labeled minimal medium:

4 g KH_2PO_4

4 g K_2HPO_4

3.5 g $Na_2HPO_4 * 2\ H_2O$

1 g NaCl

1 g $^{15}NH_4Cl$

fill up to 960 ml with de-ionized water and autoclave

To these 960 ml the following were added:

10 ml $MgSO_4$ 1 M (autoclaved)

15 ml ^{13}C-Glucose 20% (sterile filtered)

2 ml trace metal stock solution (autoclaved) (for composition see Appendix C)

1 ml thiamineHCl 150 mM (sterile filtered)

1 ml ampicillin 100 mg/ml

C - trace metal stock solution (500X):

$FeSO_4 * 7\ H_2O$	4 g/l
$CaCl_2 * 2\ H_2O$	4 g/l
$AlCl_3 * 6\ H_2O$	1 g/l
$MnSO_4 * n\ H_2O$	1 g/l
$CoCl_2 * 6\ H_2O$	0.4 g/l
$ZnSO_4 * 7\ H_2O$	0.2 g/l
$CuCl_2 * 2\ H_2O$	0.1 g/l
H_3BO_3	0.1 g/l

D - Chemical shift table of NY2R

Amide proton and ^{15}N chemical shifts of N-Y2

Residue	**N /ppm**	**H^N /ppm**
Met 1		
Gly 2	112.1	8.59
Pro 3		
Ile 4	120.1	8.16
Gly 5	112.9	8.42
Ala 6	123.6	8.07
Glu 7	119.4	8.39
Ala 8	123.9	8.14
Asp 9	119.1	8.12
Glu 10	120.9	8.30
Asn 11	118.7	8.37
Gln 12	120.3	8.15
Thr 13	115.7	8.15
Val 14	122.2	8.09
Glu 15	123.8	8.33
Glu 16	121.5	8.26
Met 17	121.1	8.22
Lys 18	122.8	8.18
Val 19	121.0	8.02
Glu 20	123.9	8.37
Gln 21	120.6	8.15
Tyr 22	120.4	8.12
Gly 23	109.8	8.06
Pro 24		
Gln 25	119.8	8.49
Thr 26	114.7	8.08
Thr 27	118.5	8.09
Pro 28		
Arg 29	121.0	8.33
Gly 30	109.4	8.28
Glu 31	119.9	8.15
Leu 32	122.8	8.19
Val 33	122.6	8.04
Pro 34		
Asp 35	121.3	8.23
Pro 36		
Glu 37	121.8	8.30
Pro 38		
Glu 39	120.1	8.37
Leu 40	123.3	8.18
Ile 41	121.2	7.98
Asp 42	123.9	8.26
Ser 43	117.8	8.36
Thr 44	114.9	8.21
Lys 45	122.2	7.83
Leu 46	122.2	7.92
Ile 47	120.2	7.83
Glu 48	124.4	8.26
Val 49	121.4	8.06
Gln 50	128.5	7.89

E - chemical shift table of Blc

residue number	atom type	chemical shift /ppm
7	CA	56.2
7	CB	30.9
7	H	8.31
7	N	121
8	CA	44.8
8	H	8.35
8	N	111
9	CA	61.2
9	CB	33.4
9	H	7.84
9	N	116
10	CA	62.5
10	CB	69.3
10	H	8.35
10	N	122
11	CA	60.2
11	CB	34.3
11	H	7.55
11	N	118
12	CA	64.4
12	CB	32.4
12	H	8.02
12	N	123
13	CA	52.3
13	CB	40.4
13	H	8.32
13	N	123
14	CA	52.8
14	CB	37.3
14	H	8.11
14	N	119
15	CA	57.8
15	CB	41.2
15	H	8.99
15	N	122
16	CA	51.6
16	CB	41.5
16	H	7.47
16	N	124
17	CA	55
17	CB	16
17	H	8.93
17	N	126
18	CA	56.3
18	CB	38
18	H	7.81
18	N	111
19	CA	55.8
19	CB	28.9
19	H	7.34
19	N	115
20	CA	58.8
20	CB	42.6
20	H	7.54
20	N	121
22	CA	44.1
22	H	8.97
22	N	108
23	CA	64.6
23	CB	70.2
23	H	8.77
23	N	118
24	CA	55.8
24	CB	32.6
24	H	9.53
24	N	130
25	CA	58.3
25	CB	39.6
25	H	10.2
25	N	120
26	CA	56.1
26	CB	31.1
26	H	8.29
26	N	123
27	CA	59
27	CB	36.5
27	H	9.11
27	N	130
28	CA	50.6
28	CB	22
28	H	7.79
28	N	120
37	CA	44.9
37	H	9.07
37	N	119
38	CA	55
38	CB	43.4
38	H	7.66
38	N	116
39	CA	53.2
39	CB	38
39	H	9.21
39	N	120
40	CA	59.2
40	CB	30
40	H	8.96
40	N	117
41	CA	62.4
41	CB	33.4
41	H	6.16
41	N	116
42	CA	58.9
42	CB	73.7
42	H	9.4
42	N	117
43	CA	50.7
43	CB	21.7
43	H	9.13
43	N	125
44	CA	61.3
44	CB	71
44	H	8.37
44	N	117
45	CA	56.2
45	CB	41.7
45	H	8.84
45	N	126
46	CA	57.2
46	CB	65.7
46	H	9.92
46	N	117
47	CA	56.1
47	CB	43.7
47	H	8.88
47	N	123
48	CA	54.9
48	CB	34.4
48	H	9.14
48	N	121
49	CA	55.9
49	CB	39.9
49	H	8.68
49	N	121
50	CA	52.8
50	CB	40.1
50	H	7.6
50	N	116
51	CA	45.5
51	H	7.82
51	N	107
52	CA	44.5
52	H	7.99
52	N	107
53	CA	53.3
53	CB	45
53	H	8.95
53	N	116
54	CA	52.6
54	CB	39.7
54	H	9.1
54	N	120
55	CA	60.4
55	CB	34.1
55	H	8.22
55	N	120
56	CA	60.2
56	CB	39.7
56	H	8.47
56	N	126
57	CA	52.6
57	CB	41.2
57	H	9.05
57	N	131
58	CA	55
58	CB	37.2
58	H	9.07
58	N	127
59	CA	46.4
59	H	9.89
59	N	107
60	CA	56.9
60	CB	39.8
60	H	9.09
60	N	122
61	CA	48.9
61	CB	39.1
61	H	7.91
61	N	126
69	CA	54.5
69	CB	32.9
69	H	8.76
69	N	120
70	CA	57.4
70	CB	65.7
70	H	9.02
70	N	119
71	CA	54.7
71	CB	31.6
71	H	8.67
71	N	123
72	CA	44.3
72	H	9.88
72	N	111
73	CA	54.8
73	CB	36.6
73	H	8.24
96	N	108
97	CA	57.5
97	CB	41.6
97	H	7.6
97	N	122
98	CA	51.7
98	CB	42
98	H	9.09
98	N	124
99	CA	63.1
99	CB	29.7
99	H	8.43
99	N	124
100	CA	62
100	CB	39.4
100	H	8.55
100	N	125
101	CA	52.2
101	CB	22.3
101	H	7.6
101	N	119
102	CA	54.8
102	CB	47.5
102	H	8.18
102	N	122
103	CA	52.9
103	CB	42.1
103	H	8.41
103	N	122
104	CA	59.1
104	CB	30.2
104	H	8.67
104	N	120
105	CA	56.1
105	CB	28.9
105	H	7.96
105	N	114
108	CA	55.2
108	CB	31.2
108	H	8.79
108	N	117
109	CA	51
109	CB	23.7
109	H	9.13
109	N	122
110	CA	53.5
110	CB	45.7
110	H	8.52
110	N	124
111	CA	60.3
111	CB	36
111	H	8.87
111	N	128
112	CA	56.3
112	CB	68.2
112	H	9.77
112	N	121
113	CA	43.6
113	H	7.7
113	N	108
122	CA	52.9
122	CB	46
122	H	9.48
122	N	127
123	CA	55.7
123	CB	65.6
123	H	9.48
123	N	113
124	CA	56.9
124	CB	30.9
124	H	9.02
124	N	126
125	CA	58
125	CB	69.3
125	H	8.29
125	N	108
127	CA	58.4
127	CB	71.8
127	H	6.78
127	N	109
128	CA	59.1
128	CB	42.2
128	H	8.65
128	N	114
129	CA	56.9
129	CB	65.1
129	H	8.29
129	N	118
130	CA	66.1
130	CB	31.4
130	H	8.65
130	N	120
131	CA	60.3
131	CB	28.4
131	H	8.55
131	N	120
132	CA	66
132	CB	31.1
132	H	7.43
132	N	121
133	CA	61.3
133	CB	32.4
133	H	8.25
133	N	121
134	CA	59.2
134	CB	28.1
134	H	8.45
134	N	115
135	CA	59.3
135	CB	29.1
135	H	7.87
135	N	121
136	CA	58.2
136	CB	34
136	H	8.44
136	N	117
137	CA	58.3
137	CB	40.7
137	H	8.45
137	N	118
138	CA	55.5
138	CB	17.8
138	H	8.09
138	N	123
139	CA	66.8
139	CB	31.4
139	H	7.89
139	N	120
140	CA	55.9
140	CB	18.4
140	H	8.12
140	N	120
141	CA	67.1
141	CB	69
141	H	8.51
141	N	114
142	CA	59.6
142	CB	30
142	H	8.21
142	N	125
143	CA	56.1
143	CB	28.6
143	H	7.94
143	N	115
144	CA	46
144	H	7.84
144	N	106
145	CA	57.9
145	CB	40.7
145	H	8.25
145	N	119
152	CA	56.6
152	CB	27.1
152	H	8.55
152	N	129
153	CA	63.6
153	CB	32.9
153	H	7.69
153	N	128
154	CA	57.9
154	CB	28.5
154	H	9.04
154	N	127
155	CA	52.9
155	CB	31.1
155	H	8.11
155	N	124
157	CA	45.2
157	H	9.39
157	N	109
158	CA	58.7
158	CB	64
158	H	8.1
158	N	115
159	CA	52.6
159	CB	18.9
159	H	9.24
159	N	125
160	CA	56.8
160	CB	29.5
160	H	7.79
160	N	119
161	CA	58.2
161	CB	63.8
161	H	7.69
161	N	116
162	CA	53.3
162	CB	28.6
162	H	8.11
162	N	120
164	CA	56.2
164	CB	29.3
164	H	8.4
164	N	120
165	CA	57.3
165	CB	39.5
165	H	8.05
165	N	120
166	CA	56.3
166	CB	30.1
166	H	8.15
166	N	122
167	CA	57.7
167	CB	33.6
167	H	7.81
167	N	127

Table A1: *Assigned chemical shifts of Blc. Spectra were recorded on a 1 mM $^{13}C,^{15}N$-Blc sample in 20 mM d_4-acetate, pH 4.5, 10% D_2O, 0.5 mM TMSP at 37 °C.*

F - Plasmid maps of the expressed region of pBlc3 and pBlc5

pBlc3:
XbaI restriction site (t'ctaga)
OmpA signal sequence (underlined)
Blc (bold)
StrepII tag (bold underlined)
HindIII restriction site (a'agctt)

```
ccatcgaatggccagatgattaattcctaatttttgttgacactctatcattgatagagt
 P  S  N  G  Q  M  I  N  S  -  F  L  L  T  L  Y  H  -  -  S
tattttaccactccctatcagtgatagagaaaagtgaaatgaatagttcgacaaaaatct
 Y  F  T  T  P  Y  Q  -  -  R  K  V  K  -  I  V  R  Q  K  S
agataacgagggcaaaaaatgaaaaagacagctatcgcgattgcagtggcactggctggt
 R  -  R  G  Q  K  M  K  K  T  A  I  A  I  A  V  A  L  A  G
ttcgctaccgtagcgcaggccgccagttctcctacgccgccgcgtggcgtgaccgtagta
 F  A  T  V  A  Q  A  A  S  S  P  T  P  P  R  G  V  T  V  V
aataatttcgacgccaaccattatcttggtacctggtatgagattgcccgttttgatcac
 N  N  F  D  A  N  H  Y  L  G  T  W  Y  E  I  A  R  F  D  H
cgctttgaacgtggactggaaaaagtcaccgcaacatacagcctgcgtgatgacggcggc
 R  F  E  R  G  L  E  K  V  T  A  T  Y  S  L  R  D  D  G  G
ctgaatgtcattaataaaggctataaccctgacagaggaatgtggcagcagagtgaaggg
 L  N  V  I  N  K  G  Y  N  P  D  R  G  M  W  Q  Q  S  E  G
aaagcgtactttaccggcgcaccaactcgcgctgcgctgaaagtgtcattctttggtcct
 K  A  Y  F  T  G  A  P  T  R  A  A  L  K  V  S  F  F  G  P
ttctatggcggttataacgttattgcactcgatcgggaataccgccatgcgctggttagc
 F  Y  G  G  Y  N  V  I  A  L  D  R  E  Y  R  H  A  L  V  S
gggccggaccgcgactacctgtggatactctcccgcacgccaaccatttctgtggaagtg
 G  P  D  R  D  Y  L  W  I  L  S  R  T  P  T  I  S  V  E  V
aaacaggagatgctggcagtcgcgacccgggaagggtttgatgtcagtaaatttatttgg
 K  Q  E  M  L  A  V  A  T  R  E  G  F  D  V  S  K  F  I  W
gtacagcagcctggtagcgcttggtctcacccgcagttcgaaaaataataagcttgacct
 V  Q  Q  P  G  S  A  W  S  H  P  Q  F  E  K  -  -  A  -  P
```

pBlc5:

XbaI restriction site (t'ctaga)
Blc (bold)
StrepII tag (bold underlined)
HindIII restriction site (a'agctt)

```
ccatcgaatggccagatgattaattcctaatttttgttgacactctatcattgatagagt
 P  S  N  G  Q  M  I  N  S  -  F  L  L  T  L  Y  H  -  -  S
tattttaccactccctatcagtgatagagaaaagtgaaatgaatagttcgacaaaaatct
 Y  F  T  T  P  Y  Q  -  -  R  K  V  K  -  I  V  R  Q  K  S
agataacgagggcaaaaaatgagttctcctacgccgccgcgtggcgtgaccgtagtaaat
 R  -  R  G  Q  K  M  S  S  P  T  P  P  R  G  V  T  V  V  N
aatttcgacgccaaccattatcttggtacctggtatgagattgcccgttttgatcaccgc
 N  F  D  A  N  H  Y  L  G  T  W  Y  E  I  A  R  F  D  H  R
tttgaacgtggactggaaaaagtcaccgcaacatacagcctgcgtgatgacggcggcctg
 F  E  R  G  L  E  K  V  T  A  T  Y  S  L  R  D  D  G  G  L
aatgtcattaataaaggctataaccctgacagaggaatgtggcagcagagtgaagggaaa
 N  V  I  N  K  G  Y  N  P  D  R  G  M  W  Q  Q  S  E  G  K
gcgtactttaccggcgcaccaactcgcgctgcgctgaaagtgtcattctttggtcctttc
 A  Y  F  T  G  A  P  T  R  A  A  L  K  V  S  F  F  G  P  F
tatggcggttataacgttattgcactcgatcgggaataccgccatgcgctggttagcggg
 Y  G  G  Y  N  V  I  A  L  D  R  E  Y  R  H  A  L  V  S  G
ccggaccgcgactacctgtggatactctcccgcacgccaaccatttctgtggaagtgaaa
 P  D  R  D  Y  L  W  I  L  S  R  T  P  T  I  S  V  E  V  K
caggagatgctggcagtcgcgacccgggaagggtttgatgtcagtaaatttatttgggta
 Q  E  M  L  A  V  A  T  R  E  G  F  D  V  S  K  F  I  W  V
cagcagcctggtagcgcttggtctcacccgcagttcgaaaaataataagcttgacctgtg
 Q  Q  P  G  S  A  W  S  H  P  Q  F  E  K  -  -  A  -  P  V
```

G - Deletion and insertion primers for the generation of the Blc derivatives carrying Y1-receptor loop sequences

Deletion	**Forward 5'->3'**
d1(dI)	cctggtatgagattgcccgtgaaaaagtcaccgcaac
d2(D1)	ccgcaacatacagcctgctgaatgtcattaataaaggc
d3(dII)	cattaataaaggctataacatgtggcagcagagtgaaggg
d4(D2)	gggaaagcgtactttgcgctgaaagtgtcattc
d5(dIII)	gcgctgaaagtgtcattctttttctatggcggttataacg
d6(D3)	ggttataacgttattgcactccatgcgctggttagcggg
d7(dIV)	ccgccatgcgctggttagcctgtggatactctcccgc
Insertion	**Forward 5'->3'**
d1Y1e3C294S	ctggtatgagattgcccgttcgactggaaccaccagatcatcgcgacctctaaccacaacgaaaaagtcaccgcaac
d2Y1e3C294S	ccgcaacatacagcctgttcgactggaaccaccagatcatcgcgacctctaaccacaacctgaatgtcattaataaaggc
d3Y1e3C294S	cattaataaaggctataacttcgactggaaccaccagatcatcgcgacctctaaccacaacatgtggcagcagagtgaaggg
d4Y1e3C294S	gggaaagcgtacttttcgactggaaccaccagatcatcgcgacctctaaccacaacgcgctgaaagtgtcattc
d5Y1e3C294S	gcgctgaaagtgtcattcttttcgactggaaccaccagatcatcgcgacctctaaccacaacttctatggcggttataacg
d6Y1e3C294S	ggttataacgttattgcactcttcgactggaaccaccagatcatcgcgacctctaaccacaaccatgcgctggttagcgggccg
d7Y1e3C294S	gccatgcgctggttagcttcgactggaaccaccagatcatcgcgacctctaaccacaacctgtggatactctcccg
D71Y1e3C294S	tacagcctgcgtgatttcgactggaaccaccagatcatcgcgacctctaaccacaacgacggcggcctgaatgtc
A101Y1e3C294S	gcgtactttaccggcgcattcgactggaaccaccagatcatcgcgacctctaaccacaaccaactcgcgctgcgctg
R126Y1e3C294S	gttattgcactcgatcggttcgactggaaccaccagatcatcgcgacctctaaccacaacgaataccgccatgcgctg
d1Y1e3m7GL	ctggtatgagattgcccgtttcgactggaaccaccagatcggactggaaaaagtcaccgcaac
d1Y1e3m5	ctggtatgagattgcccgtttcgactggaaccaccagatcatcgcggaaaaagtcaccgcaac
P133Y1e3C294S	gctggttagcgggccgttcgactggaaccaccagatcatcgcgacctctaaccacaacgaccgcgactacctgt
F53Y1e3C294S	ttttgatcaccgcttttcgactggaaccaccagatcatcgcgacctctaaccacaacgaacgtggactggaaa
d1Y1e1	ctggtatgagattgcccgttacaccctgatggaccactgggttttcggtgaagcgatggaaaaagtcaccgcaac
d3Y1e1	cattaataaaggctataactacaccctgatggaccactgggttttcggtgaagcgatgatgtggcagcagagtgaaggg
d5Y1e1	gcgctgaaagtgtcattcttttacaccctgatggaccactgggttttcggtgaagcgatgttctatggcggttataacg
d7Y1e1	gccatgcgctggttagctacaccctgatggaccactgggttttcggtgaagcgatgctgtggatactctcccg
d1Y1e2C197S	ctggtatgagattgcccgtcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttc ttttgatcagttcccgtccgattctcaccgtctgtctgaaaaagtcaccgcaac
d3Y1e2C197S	cattaataaaggctataaccaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttc ttttgatcagttcccgtccgattctcaccgtctgtctatgtggcagcagagtgaaggg
d5Y1e2C197S	gcgctgaaagtgtcattctttcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgttt cttttgatcagttcccgtccgattctcaccgtctgtcttctatggcggttataacg
d7Y1e2C197S	gccatgcgctggttagccaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttctt ttgatcagttcccgtccgattctcaccgtctgtctctgtggatactctcccg

Table A2: *Forward primers for generating the deletions and insertions in the Blc construct. Reverse primers are the complement sequences of the primers shown here. Underlined sequences are from Blc, non-underlined from the human Y1-receptor gene.*

Chapter IV: Grafting of extracellular loops of Y-receptors onto a membrane-embedded β-barrel scaffold

Abstract

In this chapter I present the attempt at creating a membrane-embedded model system, that is capable of mimicking ligand-binding functions of a subfamily of the G-protein coupled receptors (GPCRs). These receptors function in a modular way, in which a certain part of the molecule is responsible for a specific task. In our case one part is responsible for ligand-binding, whereas other parts are involved in the signal transduction process or allosteric regulation, for instance. The full receptors are large and difficult to characterize structurally, but one can resort to studying confirmed or hypothesized modules thereof. Cell surface receptors commonly are thought of consisting of three distinct modules or "domains": an extracellular, a transmembrane and an intracellular domain. We intend to study the extracellular part of a GPCR, which is responsible for binding the receptor's cognate ligand(s), by grafting it onto a stable supporting molecule, a so-called "scaffold". The resulting *chimeric protein* should be easier to produce and characterize spectroscopically than the receptor. As a potential scaffold molecule we have chosen the outer membrane protein A (OmpA) from *E. coli*, which is functionally and structurally one of the best-characterized membrane proteins.

GPCRs are of enormous clinical importance, and the availability of a suitable model for such receptors will be a valuable tool for the assessment of novel potential GPCR binding molecules.

1. Introduction

1.1 A new approach to structural and functional information on G-protein coupled receptors

G-protein coupled receptors (GPCRs) are a class of membrane embedded proteins, spanning the hydrophobic membrane interior with seven transmembrane (TM) helices, connected by three intracellular and three extracellular loops[1]. GPCRs are of enormous clinical importance[2], and as long as no straightforward approaches to the three-dimensional structures of these molecules exist, the availability of models for such receptors will continue to be a valuable tool for the investigations of many different GPCRs. One sub-class of the GPCRs is believed to interact with their cognate ligands mainly through its surface exposed extracellular parts. Our work is inspired by the idea, that these extracellular parts could be studied separately, *i.e.* in absence of the hydrophobic

transmembrane helices, which are responsible for most of the problems arising in the work with GPCRs. Several approaches to realizing this idea have been formulated. The route we are following can be called “grafting approach” and can be briefly summarized as follows: If a protein exists, which possesses a core determining its global fold and some surface exposed loops unimportant to that fold (*i.e.* a so-called “scaffold” protein), the external parts of a GPCR could be transferred to that stable core domain, without disturbing the overall global fold of the scaffold. The resulting chimeric protein could then be considered a “chimeric receptor”, because it carries all the parts of the receptor hypothesized to be important for ligand binding (hence the term “receptor”), but at the same time the geometries of the grafted loops relative to each other are determined by the stable core of the scaffold molecule. This idea is exemplified in figure 1.

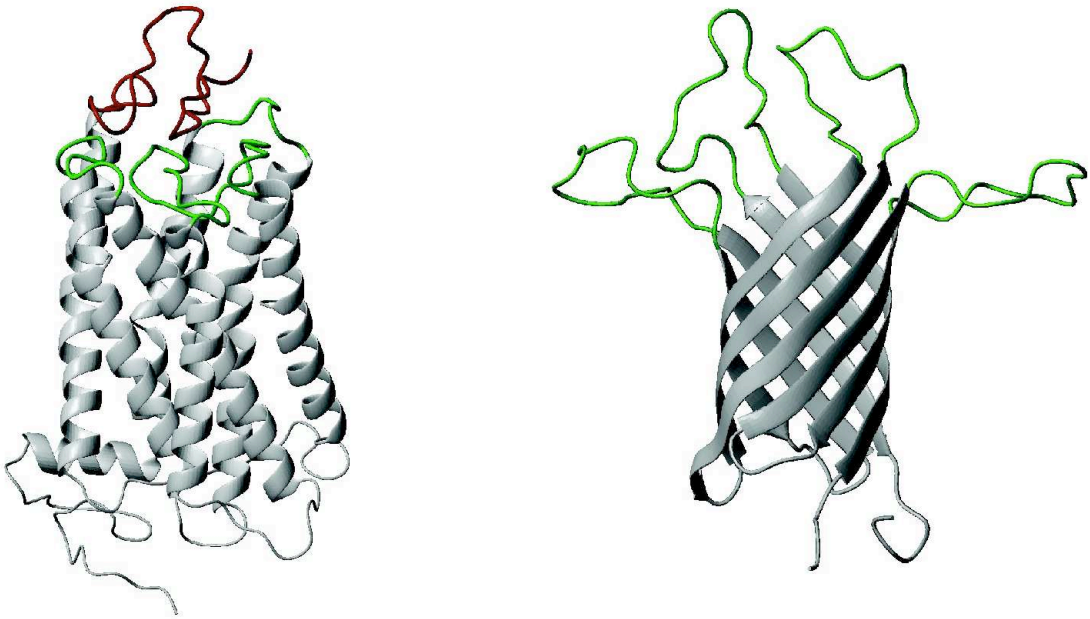


Figure 1: *Left: Structure of the prototypical GPCR bovine rhodopsin with extracellular loops and N-terminus color-coded in red and green, respectively. Right: Structure of the E. coli outer membrane protein A, with the extracellular loops colored green.*

In this work we describe the attempt at grafting the extracellular loops of a GPCR onto the membrane-embedded β-barrel of a derivative of the outer membrane protein A (OmpA) from *E. coli*. By NMR spectroscopic methods we are going to evaluate the ability of the obtained chimera to assume a stable tertiary fold, and investigate, whether these interact with the GPCR's natural ligands.

1.2 Scaffold molecules in nature and their use in synthetic protein chemistry

The advent of elaborate nucleotide synthesis and PCR schemes and highly efficient cloning techniques made it theoretically possible to produce large numbers of randomized derivatives of any given starting sequence. The theoretical number of different sequences of length n that can be composed of the 20 natural, proteinogenic amino acids rises with n in a factorial way. Therefore, already for relatively short sequences a manual expression and purification of each subsequence is practically not feasible. Through the invention of the "phage display" technology it became possible to screen large libraries for desired properties, thereby preventing the necessity to manually express, purify and test each potential candidate protein. The central idea of this technique is to create an as large as possible protein "library", where each protein and the genetic information that codes for it are physically linked. Provided an efficient selection strategy is present, one target protein along with its DNA-sequence can thus be pulled out of a mixture of billions of different proteins. Most commonly such a selection strategy consists of the surface immobilization of a source-protein and screening for target-proteins showing affinity to the source. With this technique the size of a "library" is no longer dictated by the number of sequences, which can be manually expressed and purified, but by other factors such as the transformation efficiency.

Such libraries of randomized sequences are very useful, and binders to a wide variety of different molecules can commonly be selected from one such given library. The mammalian immune system can be considered an elaborate *in vivo* protein library. The concept of generating libraries with structurally and functionally related compounds has, however, not evolved in mammals, but evolutionarily much earlier in the class of the arachnids[3].

In vitro such libraries are generated by the application of randomizing PCR techniques. The bottleneck for the generation of large numbers of randomized sequences is generally the achievable transformation efficiency, *i.e.* how many transformants one can get from a given amount of randomized DNA. This renders the randomization of the complete sequence of a typical small protein not feasible. However this is often not desired. Instead, it can be more desirable to randomize only short surface-exposed sequences. The advantages of this approach are twofold: first, the number of theoretically possible variants is much smaller, and second the unchanged sequence can, in some instances, be considered to not undergo any structural changes. If the second criterion is fulfilled by a protein one speaks of it as a "scaffold". Such a scaffold is for instance formed in the immune system by the constant regions within the immunoglobulins, or in the above mentioned arachnid peptide toxins by unchanged cysteine residues, which – by forming intramolecular disulfide bond patterns – determine the global fold of the peptides[3].

One can thus define certain characteristic features that a protein, which could possibly act as a scaffold, should present. First and foremost such a candidate scaffold should show high sequence

diversity in conjunction with a conserved tertiary fold. In other words it should have been evolved by nature to be tolerant to changes in its amino acid sequence without losing its overall folding pattern. Additionally the protein family to which the potential scaffold belongs should possess a well defined and structurally conserved (hydrophobic) core and a solvent accessible region, which is spatially and structurally well separated from the (hydrophobic) core. Ideally a potential scaffold candidate should already show diverse biochemical function among its family members. In other words nature should already have proven, that it is capable of recognizing ligands of great diversity.

1.3 The outer membrane protein A of *Escherichia coli*

The outer membrane protein A (OmpA) from *E. coli* is the most abundant protein component of the bacterial outer membrane[4] (OM) and consists of a 171 residue transmembrane and a 154 residue periplasmic domain[5]. The TM domain is an 8-stranded β-barrel and the periplasmic domain is globular. Both the N- and the C-terminus are located on the periplasmic side of the membrane[5]. The TM domain of OmpA could be expressed and purified independently of the periplasmic domain, and its structure has been solved by X-ray crystallography[6,7] and NMR spectroscopy[8,9] (figure 2a).

The β-barrel membrane anchor is the characteristic architecture of bacterial and mitochondrial outer membrane proteins (OMPs)[4]. The number of β-strands forming the barrel varies from 8 to 22 and the quaternary structure of OMPs is predominantly monomeric[10]. With one exception[11,12], all β-barrel structures solved so far display an even number of β-strands with the N- and C-termini located on the periplasmic side[10].

Electrostatically OmpA - like most β-barrel membrane proteins - is characterized by a hydrophobic barrel exterior facing the lipids of the membrane and a hydrophilic barrel interior (figure 2c). *I.e.* this membrane protein family shows an inverted electrostatic profile compared to soluble proteins. The loops interconnecting the β-strands of the barrel are hydrophilic and short on the periplasmic and long on the extracellular side[13] (figure 2a). As is the case for α-helices, the antiparallel β-sheet structure fulfills the requirement of saturating all hydrogen bonding capabilities of the backbone's carbonyl oxygens and amide hydrogens within the low-dielectric environment of the membrane[14].

The TM β-strands of β-barrel membrane proteins are rich in glycines, tryptophanes and tyrosines. The two aromatic members of this trio are mainly found in the interface regions of the membrane (figure 2b). The lumen of the smaller barrels (*i.e.* the eight-stranded OmpA) is tightly packed with an interaction network of polar residues' sidechains (figure 2c), whereas the membrane exposed surface of the barrel is hydrophobic (figure 2d).

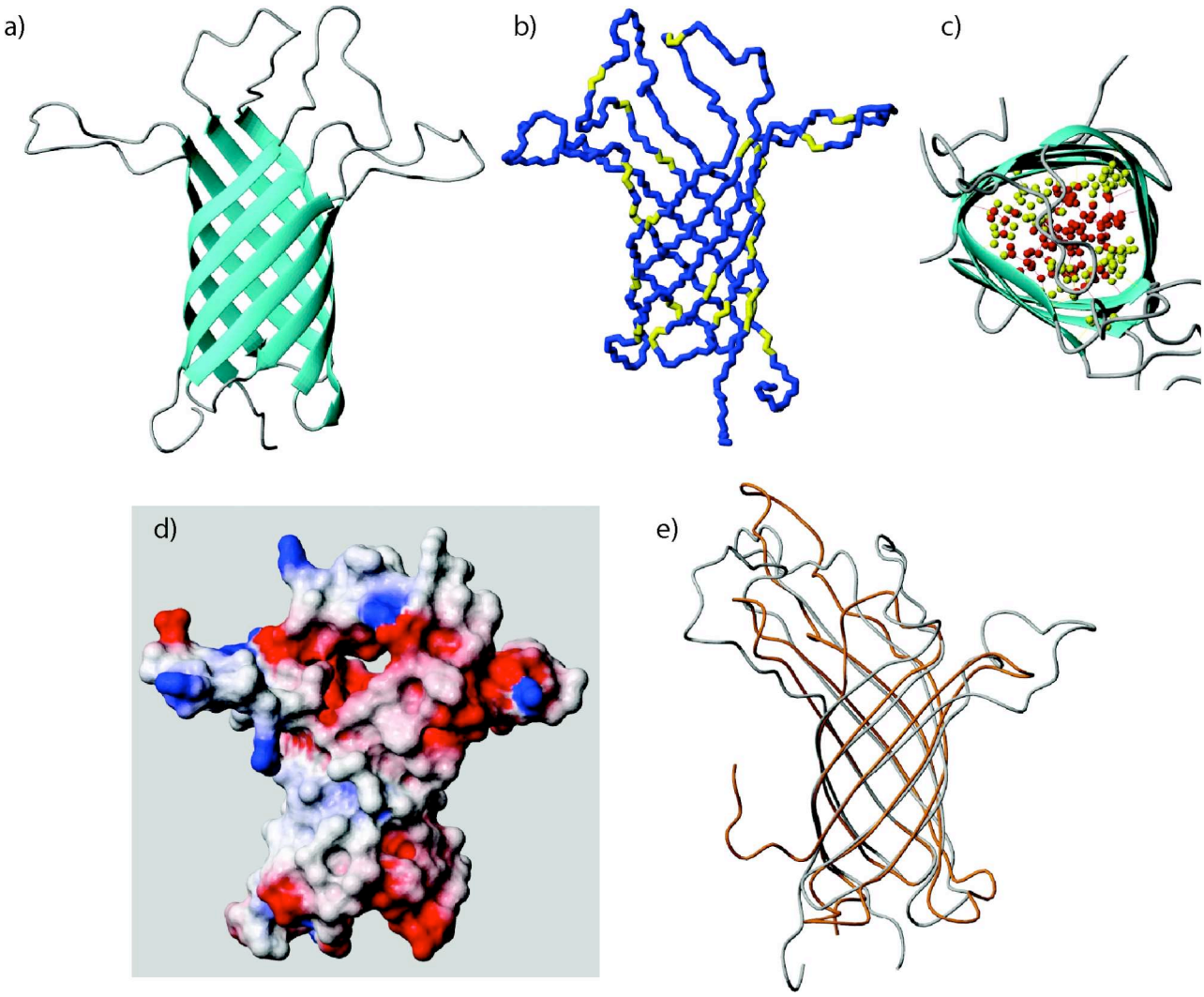


Figure 2: *a) Ribbon representation of a selected conformer of a solution NMR structure of OmpA (1G90.pdb). The β-strands of the barrel are colored cyan and delimit the surface of OmpA being inserted in the membrane. b) The characteristic aromatic belt observed in many β-barrel membrane proteins is indicated by coloring the backbone bonds of all aromatic residues of OmpA in yellow. c) View from the extracellular space into the barrel of OmpA. Charged and polar residues are represented in red and yellow, respectively. d) Electrostatic surface potential of OmpA. Negatively and positively charged residues are colored and blue, respectively. The outer surface of the central, membrane embedded region is largely hydrophobic. e) Superposition of the backbone of OmpA (gray; pdb code 2GE4) and a minimal length OmpA carrying the calmodulin EF-hand loop (orange; 2JMM).*

The surface exposed loops of OmpA have many functions. They are involved in the recognition of nutrients such as iron-siderophore complexes and sugars[15,16], toxic agents such as bacteriophages or colicins[17-19] and probably also in the recognition of eukaryotic targets for bacterial pathogens[20-22]. Denatured OmpA can be successfully refolded into detergent micelles[23] and lipid bilayers[24] by rapid dilution of the denaturant. The simplicity of OmpA refolding led to the development of expression systems, for which OmpA accumulates as insoluble inclusion bodies[25]. The inclusion body expression system allows purifying denatured OmpA in large quantities (>100 mg per liter of

bacterial culture). OmpA has long been a popular model system for the study of membrane protein (re)folding. Apart from its relative abundance in the *E. coli* OM the possibility for easily monitoring its folding state by gel electrophoresis[26] has mainly accounted for its popularity.

Extensive mutagenesis data showed the robustness of OmpA and its tolerance towards amino acid substitutions, especially in the longer extracellular loops, the short periplasmic turns and the outer face of the barrel (Koebnik, J Mol Biol, 285, 1801, 1999). The TM domain of OmpA could be circularly permutated without impairing the assembly and function in the OM (Koebnik, J Mol Biol, 250, 617, 1995). The extracellular loops could be shortened to a minimal length also without impairing the assembly and function in the OM[27]. This minimal-length OmpA has thereupon been proposed as a scaffold molecule and has been named β-barrel platform (BBP), for its potential of displaying foreign epitopes on a molecular surface of defined geometry. The grafting of a 12 residue Ca^{2+}-binding EF-hand sequence of calmodulin from *Xenopus laevis* in the position of the third extracellular loop (L3) of OmpA onto the BBP was a first successful exploitation of this minimal-length OmpA as a scaffold[28]. The BBP carrying the EF-hand sequence could be refolded in DHPC micelles and its structure was solved by NMR spectroscopy (figure 2e).

1.4 The receptors of the NPY family neurohormones: the Y-receptors

Our group has studied a wide variety of neurohormones of the neuropeptide Y (NPY) family using high-resolution NMR[29-32] and has proposed a model for the binding of these hormones to their cognate receptors, that includes membrane-association as an initial step[33]. Structural and mechanistic insight into such binding processes is anticipated to provide a basis both for the rational design of novel, and the improvement of already existing receptor-agonists and antagonists. Because of the detailed studies that have been conducted in our group on the above mentioned neurohormones and the detailed biophysical and biochemical characterization in other groups[34-37], the receptors for the NPY-family neuropeptides (Y-receptors) have been chosen as a model system for the investigation of ligand binding to GPCRs.

The members of the NPY family are C-terminally amidated polypeptides comprised of 36 amino acids and include, besides NPY, the peptide YY (PYY) and the pancreatic polypeptide (PP). Their postulated role in the regulation of food-uptake[38,39] has stirred vigorous research in many pharmaceutical companies. Seven positions are absolutely conserved among all species of NPY, PYY and PP. These are Pro5, Pro8, Gly9, Ala12, Tyr27, Arg33 and Arg35[40] (figure 3). The first atomic resolution structure of a member of the NPY family was obtained for avian PP (aPP) by X-ray crystallography[41] and later for bovine PP (bPP) by solution NMR[42]. In these structures residues 1 to 8 form a type-II polyproline helix, which is followed by a type-I β-turn connecting to residues 15 to 32, which form a α-helix, and the four most carboxy-terminal residues are flexible (figure 3).

A surprisingly stable helical hairpin is formed by backfolding of the polyproline helix onto the α-helix. This structural motif is commonly referred to as the PP-fold. The tertiary contacts, which stabilize this fold, are formed by interdigitation of the conserved proline residues in position 2, 5, and 8 on the polyproline helix and the tyrosines 20 and 27 on the α-helix[43]. The solution structure of PYY in aqueous solution was shown to be highly similar and to also display the characteristic PP-fold[32,44]. Surprisingly, in the highly homologous NPY the backfolding is absent[45,46] (figure 3). All three members of the NPY family have also been characterized in terms of their structure and dynamics in presence of dodecylphosphocholine (DPC) micelles mimicking a biological membrane (figure 3). Whereas in aqueous solution PYY and PP are structurally similar, in the presence of micelles PYY[32] and NPY[47] adopt a conformation similar to NPY, with PP taking an alternative conformation[30]. It was found that porcine NPY (pNPY) and PYY (pPYY) interact via the hydrophobic face of their C-terminal α-helix with the micelle, while the N-terminus freely diffuses in solution. bPP also interacts with the micelle via its C-terminal α-helix, but the N-terminus is also loosely associated with the micelle surface. It has been proposed, that Tyr7 mediates this association in bPP[32], which has a favorable free energy of transport into the water-membrane interface[48], in contrast to the Asn7 and Ala7 found in pNPY and pPYY, respectively. In addition also the C-terminal pentapeptide, which contains particularly important residues for receptor binding[35], differs structurally in bPP from the pNPY/pPYY pair[32].

The neuropeptides exert their effects via several receptor subtypes called Y-receptors. Four main receptors, named Y1-[49-51], Y2-[52-54], Y4-[55,56] and Y5-[57], have been cloned so far. All identified Y-receptors act via pertussis toxin-sensitive G-proteins of the G_i family. They range in size from 375 to 455 residues and show the prototypical characteristics of GPCRs of subfamily 1b, which are ligand binding primarily in the extracellular domain, and a conserved disulfide bridge between extracellular loops e1 and e2. The different receptor subtypes are localized in various tissues, both in the central nervous system and in the periphery. NPY and PYY bind equally well to the receptors Y1, Y2, Y4 and Y5 (nanomolar to sub-nanomolar dissociation constants). Only PP shows selectivity towards the Y4-receptor (picomolar dissociation constant). The ability of NPY and PYY to bind to four different receptor subtypes may be related to their conformational flexibility, which enables the peptides to adopt more than one energetically favorable conformation[58], or to the fact that they both adopt a conformation that is compatible at all subtypes.

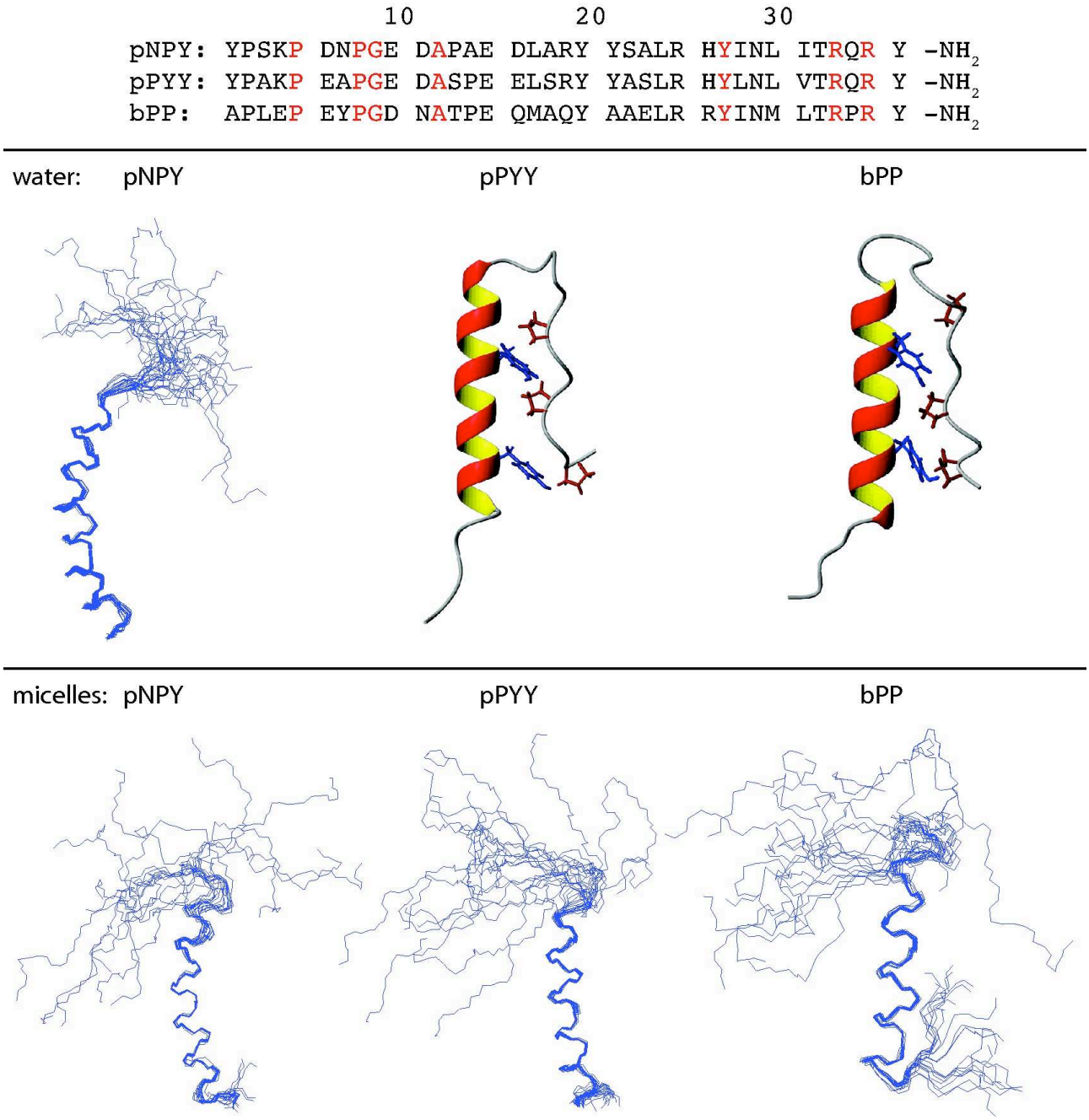


Figure 3: *Neurohormones of the NPY family. Top: Sequence comparison of porcine NPY (pNPY) and PYY (pPYY) with bovine PP (bPP). Residues, which are conserved among all three neurohormones throughout all species, are colored red. Bottom: Three-dimensional structures of pNPY, pPYY and bPP in water and in DPC micelles. In the structure of pPYY and bPP in water, the interdigitating Pro and Tyr residues are colored red and blue, respectively.*

The table below gives a summary about the most important data concerning the four major Y-receptors:

Receptor	Y1	Y2	Y4	Y5
K_i NPY [nM]	0.81[a]	0.02[a]	1.9[a]	0.19[a]
K_i PYY [nM]	1.1[a]	0.01[a]	1.1[a]	0.06[a]
K_i PP [nM]	>100[a, b]	>1000[a, b]	0.04[a, b]	27[a, b]
amino acids	384	381	375	455
major occurrence	periphery[c], hypothalamus[d]	CNS[e, f, g]	intestine, pancreas[h]	hypothalamus[i]
related action	vasoconstriction, anxiolysis[c, d, j]	memory, epilepsy, secretionc[k, l]	gastro-intestinal regulation[m]	food intake[i]

Table 1: *Biochemical, biophysical and physiological characteristics of the four major families of the Y-receptors. a) McCrea, Regul. Pept., **87**, 47-58, 2000; b) Small, Proc. Natl. Acad. Sci. U S A, **94**, 11686-91, 1997; c) Wahlestedt, Med. Biol., **64**, 85-8, 1986; d) Wahlestedt, Science, **259**, 528-31, 1993; e) Gehlert, Mol. Pharmacol., **49**, 224-8, 1996; f) Gerald, J. Biol. Chem., **270**, 26758-61, 1995; g) Rose, J. Biol. Chem., **270**, 22661-4, 1995; h) Lundell, J. Biol. Chem., **270**, 29123-8, 1995; i) Gerald, Nature, **382**, 168-71, 1996; j) Grundemar, Br. J. Pharmacol., **105**, 45-50, 1992; k) Flood, Peptides, **10**, 963-6, 1989; l) Potter, Regul. Pept., **25**, 167-77, 1989; m) Schwartz, Gastroenterology, **85**, 1411-25, 1983*

Sequence comparison shows, that among these four receptors, the Y1 and Y4 are the most closely related (42% sequence homology) while Y2 and Y5 are equally distant from each other and from the Y1/Y4 pair[40,59,60].

Little is known about the structure of any of the Y-receptor subtypes, nor have high-resolution experimental data on the complex formed between these and the neurohormones been published. Nevertheless, in many photoaffinity-labeling studies the contact points between receptors and their ligands have been assigned to the extacellular face of the GPCR. In case of human NPY at the human Y1-receptor Walker *et al*. have postulated a prominent role for Asp residues of the Y1-receptor for binding[37]. In particular, an Asp residue at the interface between the 6th TM domain and the third extracellular loop is conserved in all known Y-receptor sequences (Asp$^{6.59}$ according to the Ballesteros-Weinstein numbering scheme[61]). Other potential sites of interaction were postulated to reside within the N-terminal domain and the first extracellular loop[36], such as the converved Asp$^{2.59}$. In appendix F multiple sequence alignments for the four Y-receptors from some major mammalian species are shown.

1.4.1 The Y1-receptor

The Y1-receptor has a large number of conserved residues (figure 4a). It is therefore not straightforward to identify residues that are critical for its function.

One of the most thorough ways of characterizing the interaction of a peptide ligand with its receptor is the systematic single exchange of each residue of the ligand by L-alanine. Such a study was carried out to characterize the binding of NPY to the Y1-receptor[35]. The four natural alanines in NPY were substituted by glycine. The most important residues for binding of NPY to the Y1-receptor were according to this study Pro5, Pro8, Arg19, Tyr20 and the C-terminal positions 27-36 including Tyr27, Arg33 and Arg35. Whereas for the two C-terminal Arg residues the reduction in binding affinity was on the order of 10^4 to 10^5-fold, for the other mentioned residues it was around 10^3-fold. In constrast to the Y2-receptor N-terminally truncated versions display strongly affinity in their interaction with the Y1-receptor.

The finding that three positively charged arginine residues are critical for binding of NPY to the Y1-receptor led to the theory, that the receptor-ligand interaction might be predominantly electrostatic and therefore mediated through negatively charged residues on the receptor. This hypothesis was tested by replacing negatively charged residues in the extracellular domain of the Y1-receptor with alanines[37]. These mutants were expressed transiently in HeLa cells (epithelial cells of a cervical carcinoma of a patient called Henrietta Lacks) and their ability to bind NPY was assayed by a radioligand binding assay using ^{125}I-NPY[62]. The human Y1-receptor (hY1) contains 30 negatively charged residues of which 14 are located in a putative extracellular domain: 5 in the N-terminal part, 2 in extracellular loop 1 (e1), 6 in e2 and 1 in e3.

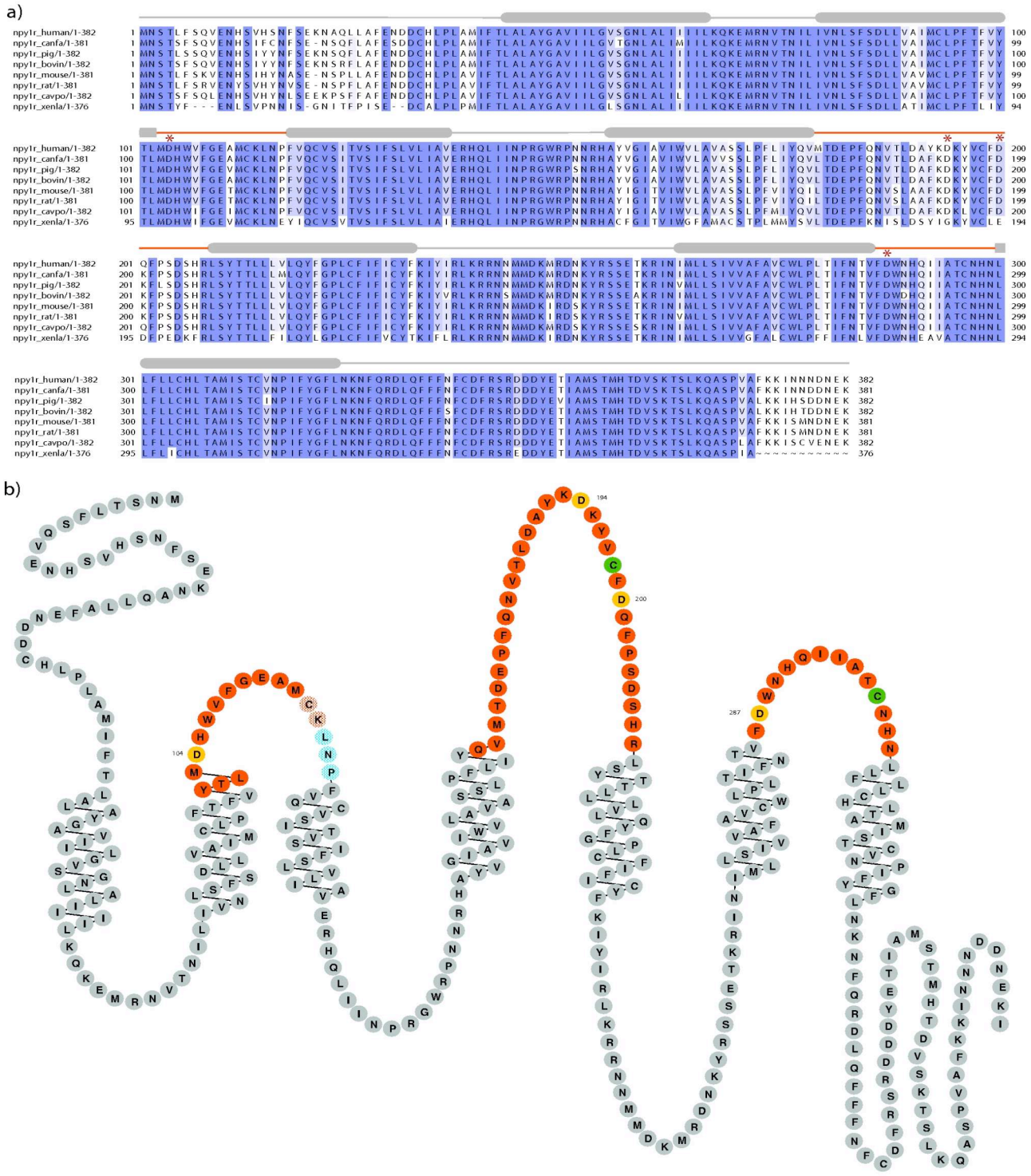


Figure 4: *a) Multiple sequence alignment of the Y1-receptor from eight major mammalian species. The intensity of the blue coloring indicates the conservation levels of different residues. Above the sequence gray rods indicate the predicted TM helices. Gray lines mark the intracellular loops and N- and C-termini. Red lines represent extracellular loops. Negatively charged residues in the extracellular loops found to be essential for ligand binding are marked with a red asterisk. b) Snake plot of the predicted topology of the human Y1-receptor. Solid orange circles indicate sequences, which have been grafted onto the scaffold. Solid yellow circles highlight negatively charged residues in the extracellular loops found to be essential for ligand binding. Solid green circles mark cysteine residues mutated to serine in the grafting constructs. The orange chessboard-patterned circles indicate residues, which have been added in a second round of grafting. The blue chessboard-patterned circles indicate residues, which are not part of the e1-loop in all predictions and have been disregarded for the grafting.*

None of the mutations introduced in the N-terminus affected NPY binding. In e1 the D104A substitution led to a complete loss of affinity for NPY, whereas the E110A mutant didn't show any altered NPY binding. In e2 two aspartates, namely Asp194 and Asp200, were essential for NPY binding and the mutation of Asp205 resulted in a significant loss of binding affinity (figure 4). The other 3 positions proved to be tolerant for substitution. The D287A mutation in the e3 again resulted in the complete loss of affinity for NPY. It is interesting to note, that the mutations affecting NPY binding the most are all clustered at the top of TM helices 5 and 6.

2. Results

2.1 Design aspects

2.1.1 Anchor point geometries of the extracellular loops in GPCRs

So far atomic structural details at atomic resolution are available for four major different GPCRs. These are rhodopsin, the β1- and β2-adrenergic receptors and the A_{2A} adenosine receptor. As of December 2009 there have been 24 GPCR structures at atomic resolution deposited in the Protein Data Bank (www.rcsb.org). 18 of these are structures on various isoforms of rhodopsin and 4 are of the β2-adrenergic receptor. A compilation of the deposited data sets can be found in Appendix E. Generally the crystallographic data show high B-factors in the cytoplasmic and in the extracellular loops (figure 5a), indicating flexibility in the loop regions, while the TM helices are more rigid. This is in accordance with experiences gained during the crystallization process of the GPCRs[63] and NMR data on the 7TM protein sensory rhodopsin pSRII[64].

We define here the term "anchor points" for the loops of the GPCRs as the C_α atoms of those TM residues lying at the border between TM helices and loops, *i.e.* the C_α atoms of the terminal residues in the TM helices. Two such anchor points form the attachment site for one cytoplasmic or extracellular loop. For the three extracellular loops of a GPCR the extracellular anchor points of helices II/III, IV/V and VI/VII form the attachment sites for extracellular loops 1 (e1), 2 (e2) and 3 (e3), respectively. The orientation of these attachment sites relative to each other defines the overall topology of the domain formed by the extracellular loops. A comparison of the distances between those anchor points on the extracellular side for a selected subset of GPCRs of known structure is presented in figure 5b. The spacing between the attachment sites for the three extracellular loops is at an average of 13 Å for e1 and e2 and 14 Å for e3 and shows a narrow distribution of roughly +/- 3 Å. The distances between the loops themselves, say between the C-terminal anchor point of the e1 and the N-terminal anchor point of the e2 loop, are less well conserved. This indicates that the

relative positions of two helices anchoring an extracellular loop are more conserved than the relative positions between helices not directly connected by an extracellular loop. The average over the distances observed for the analyzed GPCRs served as an estimate for the distances, which can be expected in the Y-receptors.

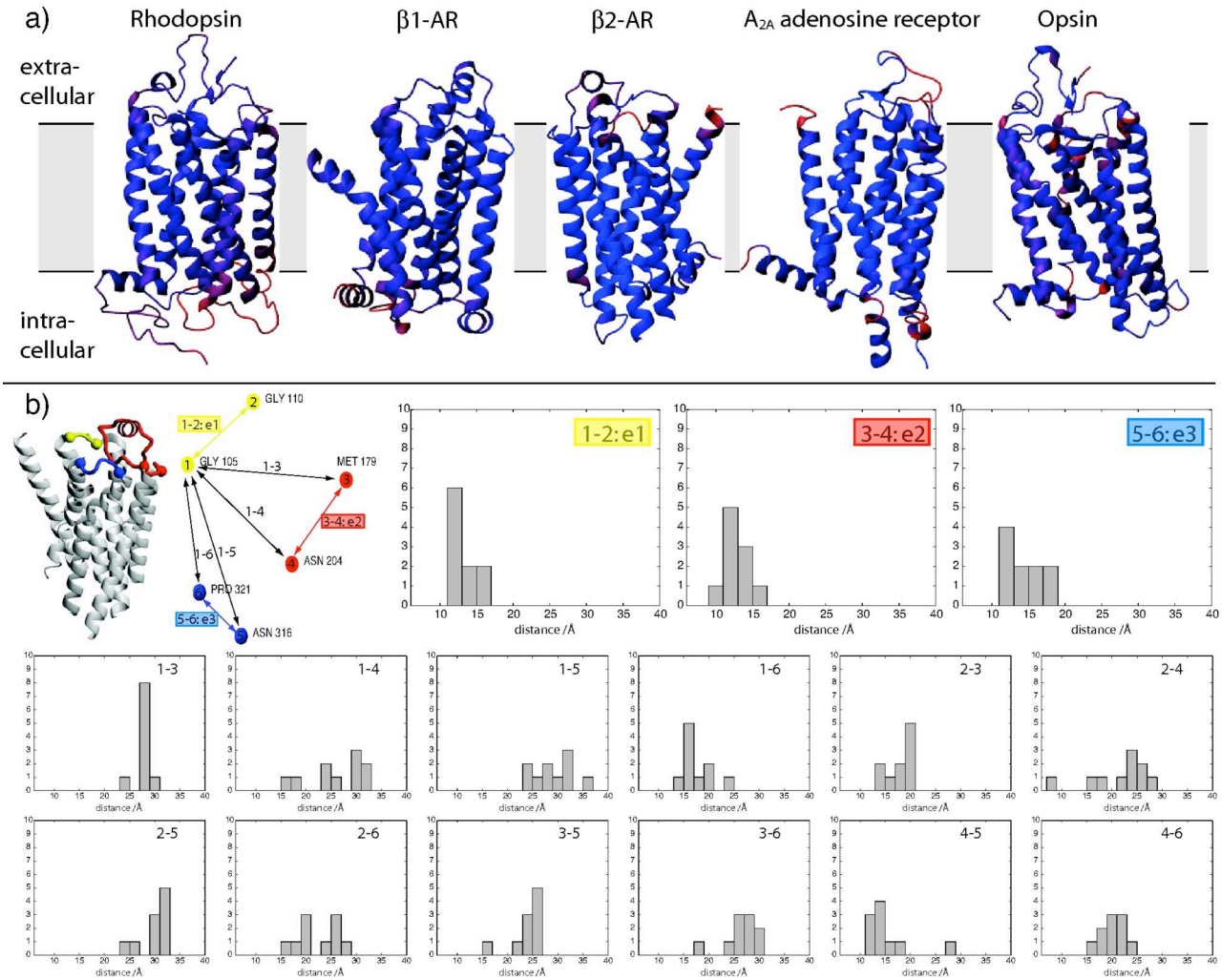


Figure 5: *a) Ribbon representation of 5 selected GPCRs. The ribbons are color coded according to the B-factor of the underlying backbone atoms. Blue stands for a low B-factor and red for a high B-factor. b) Distance statistics for the anchor points observed in a set of 10 selected GPCRs. The ribbon representation of the β1-AR has the extracellular loops colored (e1 yellow; e2 red; e3 blue).*

The Y-receptor sequences to be grafted were selected based on the predicted topology of the individual members of this family, derived from biochemical experiments and bioinformatics tools. The extracellular loops are about 15 residues long for e1 and e3 and 30 residues for e2. The predicted topology of the Y-receptors and the loops sequences chosen to be grafted are shown in figure 4b.

2.1.2 Anchor point geometries of the extracellular loops in OmpA and minimal-length OmpA

The importance of the extracellular loops of OmpA for folding has been studied by shortening those loops. A minimal-length OmpA, which had all four extracellular loops radically truncated, was found to still retain full folding capacity[27]. The sequence of this minimal-length OmpA compared to the wildtype sequence is shown in figure 6a. The ~10 residue extracellular loops of OmpA have been replaced with one or two amino acids allowing for the formation of a short β-turn.

The structure of the β-barrel TM domain of OmpA has been solved both by X-ray crystallography[6,7] and NMR spectroscopy based on NOE restraints[8] and refined using residual dipolar couplings in addition to the NOE restraints[9]. A study using the minimal-length OmpA described above as a scaffold for the Ca^{2+}-binding EF-hand loop from *X. laevis* has yielded a high resolution NMR structure as well[28] (figure 2e). Not surprisingly all five of these structures show higher flexibility of the extracellular loops as compared to the β-strands of the barrel and the short periplasmic turns. This flexibility is reflected in increased B-factors for the crystal structures and higher rmsd values for the NMR structures. This is illustrated in figure 6b, in which the structural elements are color-coded according to their B-factors or backbone rmsd values. Both crystal structures show a pronounced asymmetry of the β-barrel with the strands 4, 5 and 6 being much longer than the other five strands. This type of barrel-asymmetry is less distinct in the NMR structures (figure 6a). Additionally both crystal structures are lacking defined electron density for a substantial number of residues (37 for 1QJP and 22 for 1BXW), all located in the region of the extracellular loops. We therefore relied only on the NMR structures for the analysis of the distances and geometries of the extracellular loops and their anchor points. We use the term "anchor points" here in analogy to the way it was used for the GPCRs, *i.e.* the C_α atoms of the terminal residues in the β-strands of the barrel. In figure 6c a statistical analysis of the distances observed in the ten lowest energy conformers of the three NMR structures is presented.

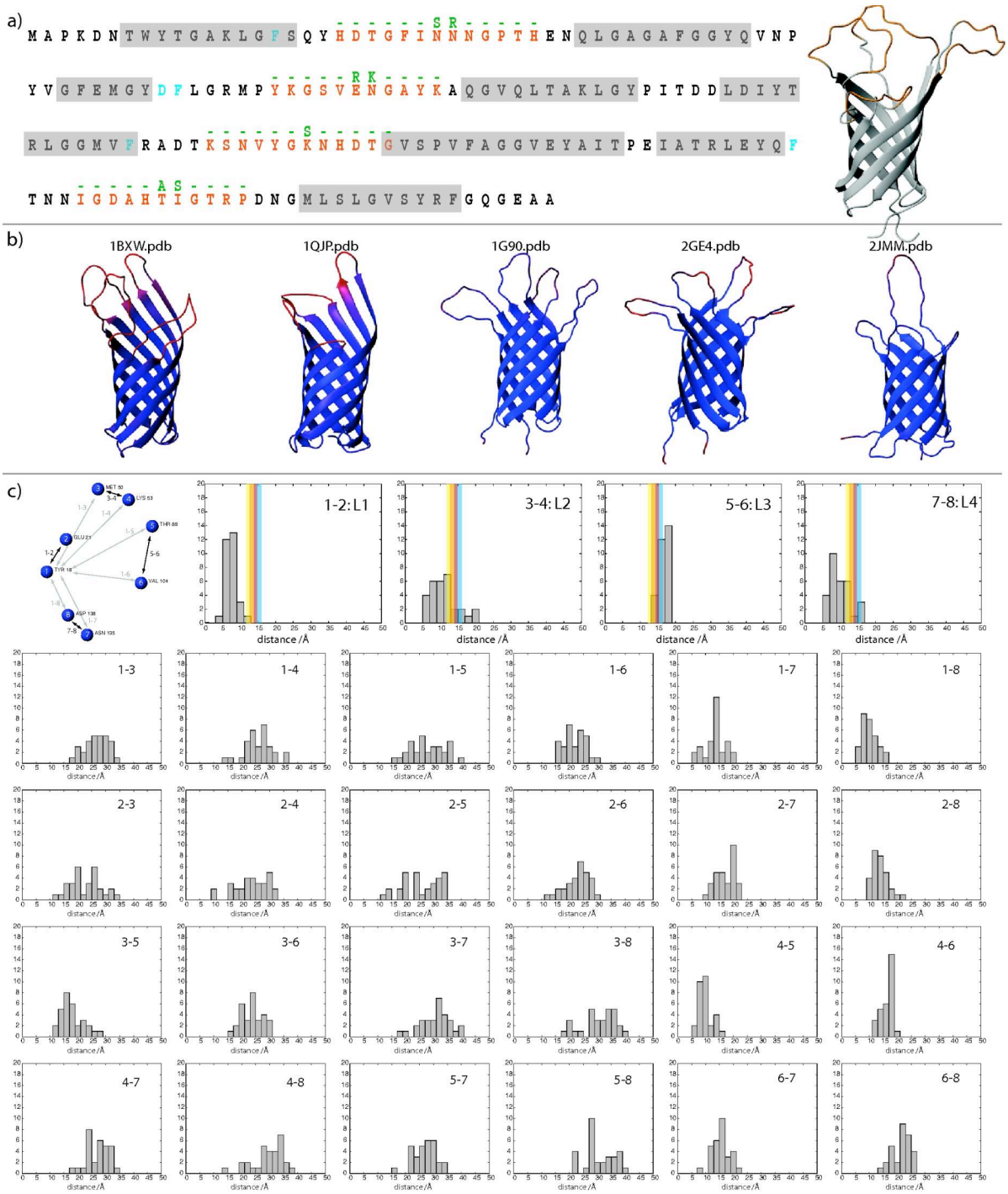


Figure 6: *a) Sequence of the OmpA construct whose structure has been solved by NMR spectroscopy (2GE4). The β-strands are indicated by gray rectangles, and the extracellular loops are colored red. Cyan letters are mutations compared to our sequence: phenylalanines are replaced with tryptophans and the aspartate with a glutamate in our construct. b) Ribbon representations of the two crystal structures (1BXW and 1QJP), the two NMR structures (1G90 and 2GE4) of OmpA and the minimal-length OmpA carrying the EF-hand loop in position of loop 3 (2JMM). The ribbons are color coded according to the B-factor of the underlying backbone atoms. Blue stands for a low B-factor and red for a high B-factor. c) Distance statistics for the anchor points of the OmpA β-barrel observed in the 3 NMR structures (1G90, 2GE4, 2JMM). The top four histograms represent the distances between the anchor points connected by the extracellular loops 1 to 4. The average distances observed between the anchor points of the extracellular loops of the GPCRs are indicated with yellow (e1), red (e2) and blue (e3). The lower 24 histograms represent the distances between all other anchor points not directly connected by an extracellular loop.*

2.1.3 Comparison of anchor points in GPCRs and the scaffold

The analysis of the distance between anchor points in GPCRs and OmpA shows, that these distances are more broadly distributed for OmpA than for the ten selected GPCRs (figures 6c and 5b). The fact that the distance distribution observed in the GPCRs falls within the distribution found in the OmpA structures, suggests that the β-barrel of OmpA might indeed be a good molecular scaffold for the extracellular domain of GPCRs.

It should be noted here, that GPCRs possess 3 and OmpA 4 extracellular loops. Therefore depending on the number of Y-receptor loops to be grafted, at least one "acceptor" site in the scaffold will remain unoccupied. In order to minimize the possibility of the natural loop residues from that site interfering with any potential ligand binding, an acceptor site, not filled with a GPCR loop, can be replaced by the minimal, turn-inducing motif found in the study of Koebnik[27], allowing the formation of a β-turn between the two strands of the barrel.

2.1.4 Construct design

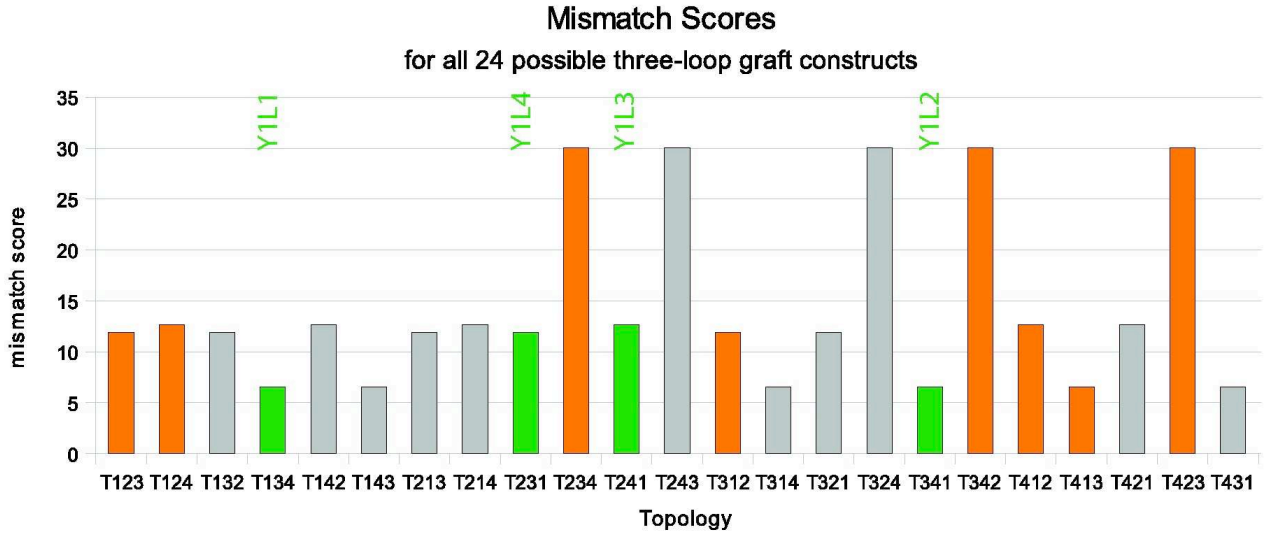
The frequencies of usage of a given codon are very different in *E. coli* and eukaryotic cells[65,66]. As the Y-receptor gene sequences were from human, we optimized the loop sequences to be grafted, in order to match with the preferred *E. coli* codon usage.

In order to determine the compatibility of the Y1-receptor loops with our scaffold, we exchanged each of the four extracellular loops of OmpA with each of the three Y1-receptor extracellular loops. The resulting 12 constructs were called "one-loop exchange constructs", since only one OmpA-loop was exchanged by a Y1-receptor loop and the other three OmpA extracellular loops were left in place in these constructs. This is in contrast to the "one-loop graft constructs", where a Y1-receptor loop was grafted into one OmpA acceptor site and the other three sites were filled with a minimal turn-inducing motif of 1-2 residues, *i.e.* one Y1-receptor loop was grafted onto the minimal-length OmpA.

Residue $D^{6.59}$ of the Y1-receptor was identified as one of the most important for high-affinity ligand binding[67]. It is located in the third extracellular loop of the Y1-receptor (Y1e3) (figure 4). We therefore grafted the Y1e3 loop (Cys294 was replaced with a serine residue in these constructs) onto all four acceptor sites of the OmpA scaffold, replacing the other three loops of OmpA with a minimal turn-inducing motif. These four constructs were named "one-loop graft constructs".

The ultimate goal of this study was to transfer all three extracellular loops of a Y-receptor simultaneously onto the minimal-length OmpA scaffold. Theoretically there are 24 different ways of arranging three donor loop sequences on the four acceptor sites of the scaffold. However, not all of these topologies are expedient. In order to rule out unpromising constructs, we calculated a "mismatch score" for each of the 24 possible arrangements. Distance mismatches between all the

relevant anchor points (*i.e.* those six which were attached to a donor loop sequence, but not those two which were connected to a minimal turn-inducing sequence) of the scaffold and those of an average GPCR were calculated and summed (figure 7). Those constructs giving the lowest mismatch scores were considered the most promising candidates. Among the group of candidates with low mismatch score, only those with a correct sequential arrangement of the loops (*i.e.* the C-terminus of e1 should be followed by the N-terminus of e2) were considered. From these we have selected four topological arrangements for the grafting of all three Y1-receptor extracellular loops onto the minimal-length OmpA scaffold simultaneously and named these constructs "three-loop graft constructs". These four constructs were given the names Y1L1, Y1L2, Y1L3 and Y1L4 (figure 7).



Y1L1: OmpA-d1Y1e1-d2-d3Y1e2C197S-d4Y1e3C297S

	d1Y1e1	d2	d3Y1e2	d4Y1e3
ΔN	0	-11	+22	+3
Δpl	-1.62	-	-4.06	-0.77
Δ -	+1	-1	+5	0
Δ +	0	-1	+1	-1

Y1L2: OmpA-d1Y1e3C297S-d2-d3Y1e1-d4Y1e2C197S

	d1Y1e3	d2	d3Y1e1	d4Y1e2
ΔN	+1	-11	+1	+23
Δpl	0	-	-4.16	-2.29
Δ -	0	-1	+1	+5
Δ +	0	-1	-2	+2

Y1L3: OmpA-d1Y1e3C297S-d2Y1e1-L3-d4Y1e2C197S

	d1Y1e3	d2Y1e1	L3	d4Y1e2
Δ N	+1	+2	-11	+23
Δ pl	0	-2.07	-	-2.29
Δ -	0	+1	-1	+5
Δ +	0	-1	-2	+2

Y1L4: OmpA-d1Y1e3C297S-d2Y1e1-d3Y1e2C197S-L4

	d1Y1e3	d2Y1e1	d3Y1e2	L4
ΔN	+1	+2	+22	-9
Δpl	0	-2.07	-4.06	-
Δ -	0	+1	+5	-1
Δ +	0	-1	+1	-1

Figure 7: *The mismatch scores of the 24 different topologies with which the three extracellular loops of the Y1-receptor can be arranged on the four acceptor sites of the minimal length OmpA scaffold. Topologies with a correct sequential arrangement of the donor loops are colored orange and green, with the latter highlighting the four expressed. The tables at the bottom compare the Y1-receptor loops and the minimal linker sequences with the native OmpA loops, which they are replacing in terms of loop length (ΔN), isoelectric point (Δpl), number of negatively and positively charged residues in the loops (Δ- and Δ+).*

2.2 Synthetic aspects

2.2.1 Expression, purification and refolding of wt-OmpA

OmpA, which has been denatured by chaotropic agents, can be successfully refolded into detergent micelles[23] and lipid bilayers[24] by rapid dilution of the denaturant. The deletion of the DNA sequence coding for the periplasmic export signal results in the accumulation of OmpA in the cytoplasm in the form of insoluble inclusion bodies. This led to the development of expression systems which allow to purify denatured OmpA in large quantities (>100 mg per liter of bacterial culture)[25]. We have overexpressed unlabeled and ^{15}N-labeled OmpA in that way and purified it according to established protocols[24,68]. ^{15}N-labeled OmpA and all derived constructs were expressed in M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source. OmpA was purified in denatured form in 8 M urea, with yields of ~200 and ~100 mg per liter of unlabeled LB- and ^{15}N-labeled M9-culture, respectively.

The folding state of many bacterial OMPs, including OmpA, can be conveniently monitored by non-denaturing SDS-PAGE[26]. The refolding behavior of OmpA was studied by diluting an urea-denatured OmpA solution into acetate (pH 4), HEPES (pH 7) or borate (pH 10) buffers containing β-dodecyldimaltoside (DDM)[69], β-octylglucoside (β-OG)[69], tetraethyleneglycol monooctylether (C_8E_4)[69,70], N-lauryldimethyl amineoxide (LDAO)[69], dodecylphosphocholine (DPC)[69,71] or dihexanoylphosphatidylcholine (DHPC)[72,73] at concentrations well above their critical micellar concentration (cmc) values and at a detergent/protein ratio >500. The table below shows the relevant biophyscial constants of these detergents, along with the concentrations that were used and the resulting detergent:protein and micelle:protein ratios.

Detergent	Charge	Molecular Weight [g/mol]	Critical micelle conc. [mM]	Aggregation number	used conc. [mM]	detergent: protein ratio	micelle:protein ratio
DDM	Nonionic	511	0.2	110-140	13	524	>3.6
β-OG	Nonionic	292	19-25	90	100	4014	>33
C_8E_4	Nonionic	306	8	82	25	1013	8.3
LDAO	Zwitterionic	229	2	69-73	14	560	>6.6
DPC	Zwitterionic	352	1.5	50-60	14	568	>8.3
DHPC	Zwitterionic	454	15	19-35	66	2643	58

Table 2: *Biophysical constants of the detergents used in our refolding screen. All values were taken from Sanders, 2006[74]. DDM: β-dodecyldimaltoside, β-OG: β-octylglucoside, C_8E_4: tetraethyleneglycol monooctylether, LDAO: N-lauryldimethyl amineoxide, DPC: dodecylphosphocholine, DHPC: dihexanoylphosphatidylcholine*

The refolding efficiency was monitored by SDS-PAGE. It increased with increasing pH (figure 8a). Incubation of the refolding reaction for 5 h at 30 °C in 10 mM borate, 1 mM EDTA, pH 10 resulted in near complete folding for the nonionic detergents DDM and β-OG and the zwitterionic DHPC and DPC. The nonionic detergent C_8E_4 and the zwitterionic LDAO resulted in only partial folding at pH 10 (figure 8a).

^{15}N-labeled OmpA was refolded at pH 10 into β-OG. Not surprisingly the resulting NMR spectrum showed less than the expected 168 peaks due to fast amide hydrogen exchange of solvent exposed loop residues at pH 10. Lowering the pH to <7 resulted in partial precipitation and a deterioration of the NMR spectrum. OmpA was finally refolded into DHPC micelles according to a refolding protocol published for the closely related OmpX[75], which yielded spectra of good quality at pH 6.5, as can be seen in figure 8b.

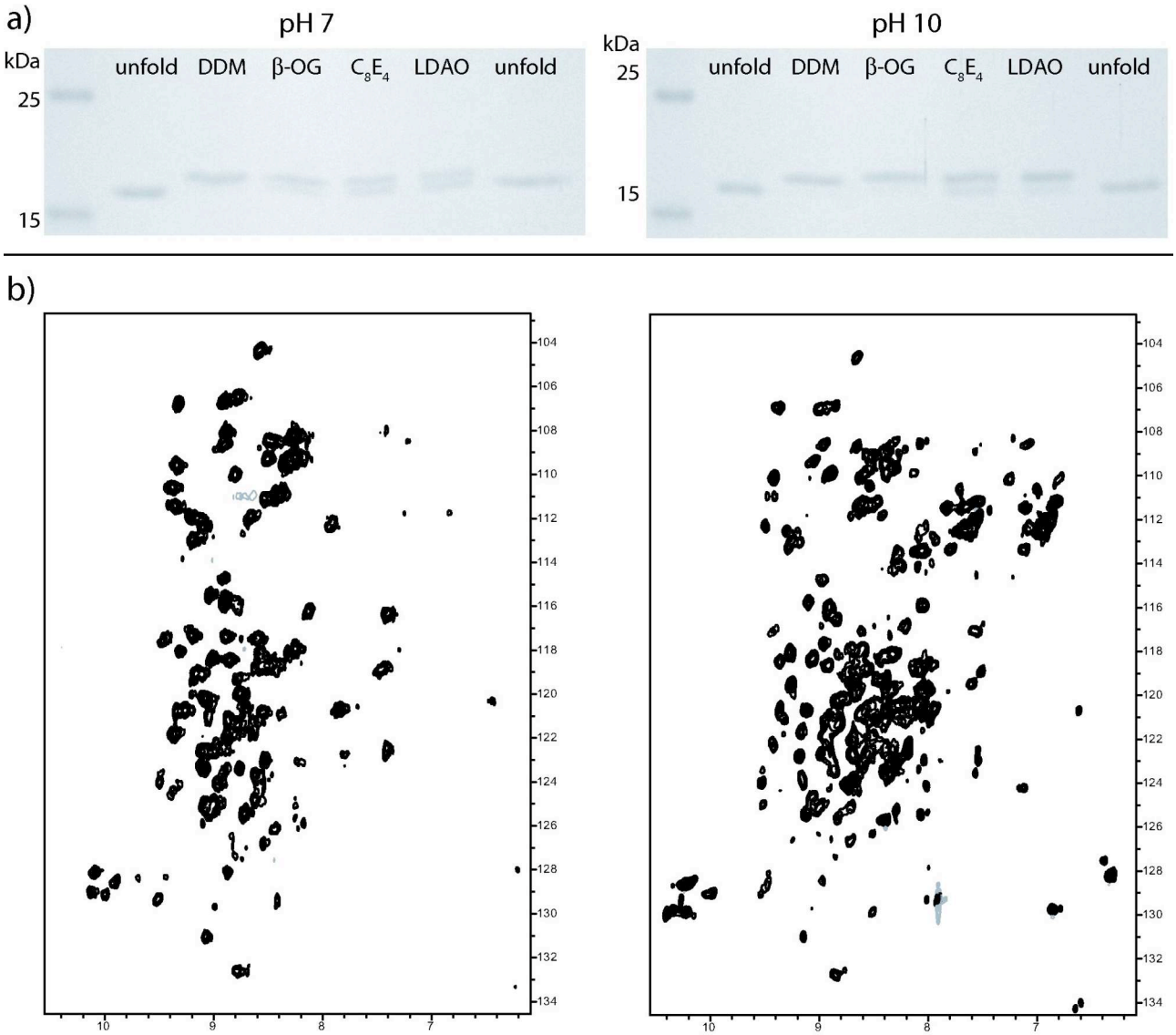


Figure 8: *a) Refolding of OmpA in four different detergents (for abbreviations and concentrations of the detergents see table 2) at pH 7 and pH 10. The denatured form of OmpA corresponds to a heat-denatured sample in presence of SDS. b) [^{15}N,^{1}H]-HSQC spectra of OmpA in β-OG at pH 10 (left spectrum) and in DHPC at pH 6.5 (right spectrum).*

2.2.2 Expression, purification and refolding of the one-loop exchange and one-loop graft constructs

All 12 possible one-loop exchange constructs were expressed and purified in unlabeled form and a selected subset in ^{15}N-labeled form with yields similar to the ones obtained for wt-OmpA. Concerning the refolding efficiency, the one-loop exchange constructs showed a similar pH dependence than observed for wt-OmpA, *i.e.* increasing efficiency with increasing pH (data not shown). In figure 9a the refolding behavior of all 12 constructs in a selected detergent (LDAO) is depicted along with a more detailed screening for a few of the constructs (figures 9b and 9c).

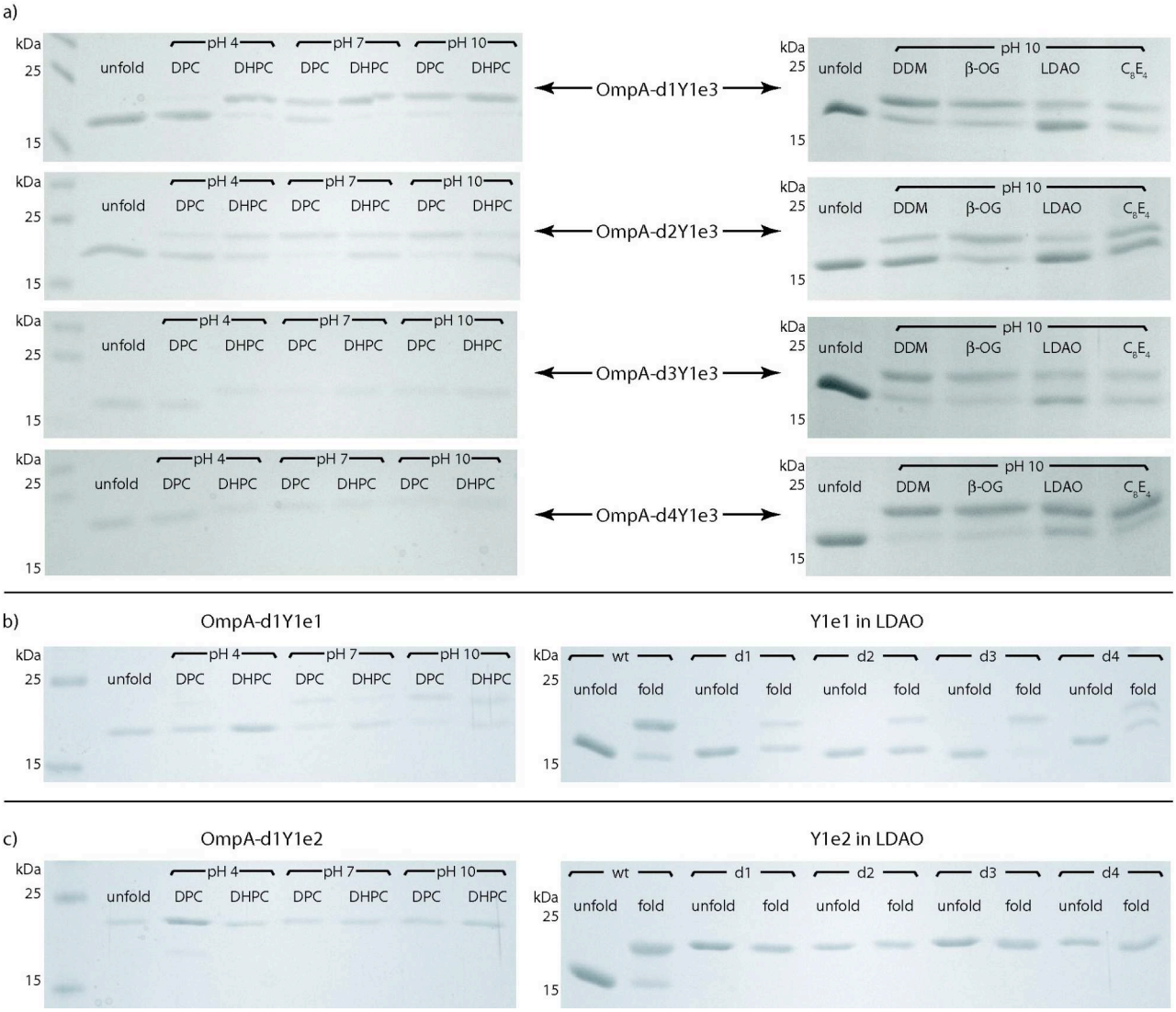


Figure 9: *Refolding behavior of all 12 one-loop exchange constructs studied by SDS-PAGE. The term "unfold" always refers to the heat-denatured of the respective species. a) Y1e3 loop exchange constructs: OmpA loop positions are numbered d1 to d4. I.e. d1Y1e3 designates, that the Y1e3 loop replaces the extracellular loop 1 of OmpA. Left column: All four constructs carrying the Y1e3 loop were refolded in two different detergents (DPC and DHPC) at four different pH values (4, 7 and 10). Right lane: Their refolding behavior in a set of detergents at the most favorable pH is illustrated. b) Y1e1 loop exchange constructs: Left: The construct carrying the Y1e1 loop in the d1 position was refolded in two different detergents (DPC and DHPC) at four different pH values (4, 7 and 10). Right: The refolding behavior of all four Y1e1 loop exchange constructs at pH 10 in LDAO is compared to the behavior of wt-OmpA. c) Analogous to b), but for the Y1e2 loop exchange constructs.*

These results demonstrate the compatibility of the Y1-receptor loop sequences with the expression and folding of the OmpA β-barrel core. Note that the clear separation of folded and bands for wt-OmpA and the constructs carrying the Y1e1 and Y1e3 loops is absent in the constructs carrying the Y1e2 loop. To test that this absence is not due to the inability of the Y1e2 constructs to adopt a stable tertiary structure, we conducted NMR experiments, which indicated through their chemical shift dispersion, that folded constructs were indeed present.

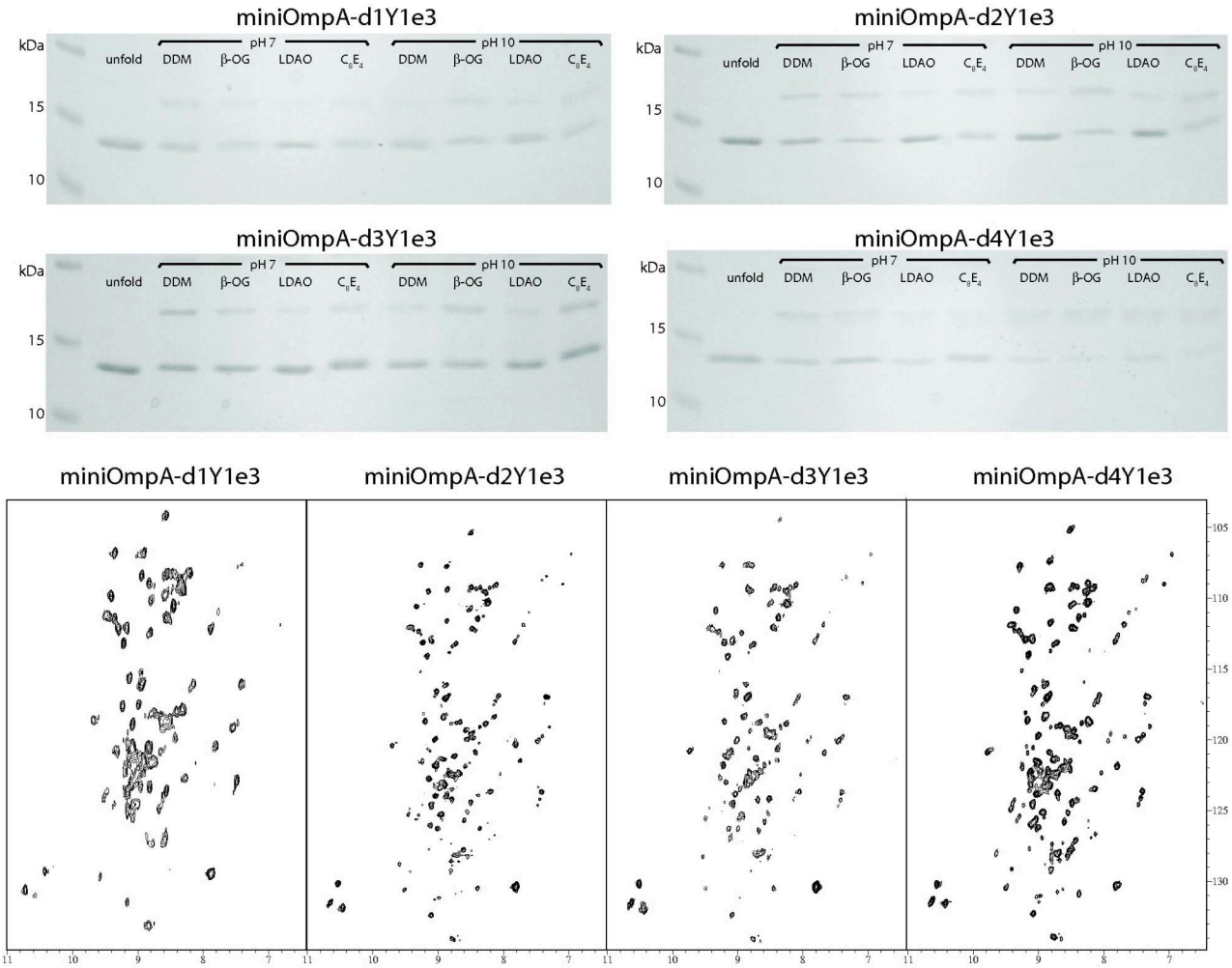


Figure 10: *Refolding efficiency of four one-loop graft constructs. The Y1e3 sequence was grafted onto the minimal length OmpA (miniOmpA) in all four acceptor positions (denoted as d1 to d4). The SDS-PAGE shows a decreased refolding efficiency compared to wt-OmpA, which can be refolded completely at pH 10. For each of the four constructs the refolding at two different pH values and four different detergents is shown. "Unfold" refers to the heat-denatured form of the respective construct. The NMR spectra were all acquired with 0.5 mM protein concentration in 10 mM borate buffer at pH 10 and in the presence of 14 mM LDAO at 47 °C.*

The four one-loop graft constructs could be expressed and purified in the same manner as the wt-OmpA and the one-loop exchange constructs with similar yields. Refolding trials with these constructs in DDM, β-OG, LDAO and C_8E_4 micelles showed less pH dependence compared to wt-OmpA, as maximum refolding yields could be achieved both at pH 7 and 10. However, these maximum yields were at most 50% (in β-OG and C_8E_4), which is significantly lower than the almost complete refolding achieved for wt-OmpA (in DDM and β-OG) (figure 8a).

A priori, a different migration behavior in SDS-PAGE of the constructs in micellar solution from heat- or chemically denatured forms of the same construct is not a proof of the assumption of a defined tertiary fold. However, the signal dispersion in the proton dimension of a protein is a good indication for the presence or absence of a defined tertiary fold. The dispersion found in the [^{15}N, ^{1}H]-HSQC spectra (figure 10) clearly indicates such a fold. It should be noted here, that the refolding of these one-loop graft constructs showed only ~50% efficiency. No attempt was made at

increasing the yields of the refolding reaction or at separating the folded from the unfolded component, as these one-loop graft constructs were only intermediates on the way to constructs carrying all three extracellular loops of the Y1-receptor (three-loop graft constructs).

2.2.3 Expression, purification and refolding of the three-loop graft constructs

The four three-loop graft constructs Y1L1, Y1L2, Y1L3 and Y1L4 could be expressed and purified with similar yields than wt-OmpA.

As the removal of the natural extracellular loops of OmpA and their replacement with the Y1-receptor loops or minimal turn-inducing motifs represents a substantial change of the overall sequence of OmpA, the ability of the constructs to assume a defined three-dimensional structure had to be assayed first. This was done in the same way as described for the one-loop exchange and one-loop graft constructs. A large set of refolding conditions were screened by SDS-PAGE in order to find the conditions giving the best yields of folded constructs. The SDS-PAGE assay allowed the assessment of a relatively large number of conditions in parallel and required only small amounts of sample (micrograms of protein) and lipids (milligrams). The most promising conditions were refolded on a larger scale (milligrams of protein and hundreds of milligrams of lipids) and subjected to NMR spectroscopic analysis. Figures 11 and 12 show the elcetrophoretic migration behavior of a selected subset of refolding conditions of the four constructs on SDS polyacrylamide gels and [^{15}N,^{1}H]-HSQC spectra acquired under the most promising conditions. Note that the migration behavior of the folded compared to the unfolded species is reversed in the four constructs when compared to the other species studied so far: While folded OmpA (figure 8), the one-loop exchange constructs (figure 9) and the one-loop graft constructs (figure 10) migrate slower than the respective unfolded speices, the three-loop graft constructs all migrate faster in the folded state than in the unfolded sate.

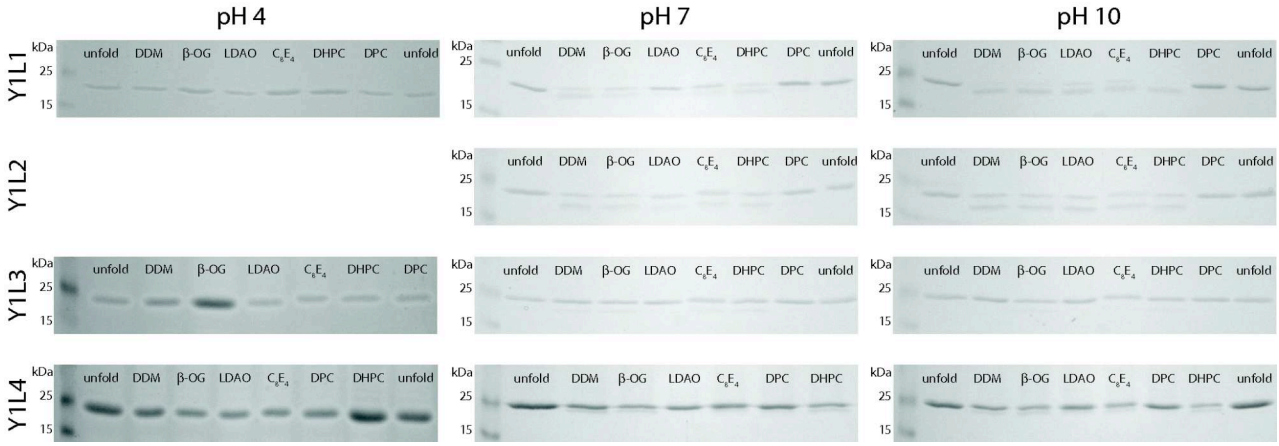


Figure 11: *SDS-PAGE analysis of the refolding behavior of the four three-loop graft constructs Y1L1, Y1L2, Y1L3 and Y1L4. Refolding efficiencies at three different pH values and in six different detergents are shown. "Unfold" refers to the heat-denatured form of the respective construct.*

A refolding screen was conducted, spanning the pH range from 4 to 10 and a variety of 6 different detergents (DDM, β-OG, LDAO and C_8E_4, DPC and DHPC) at concentrations well above their cmc values and at a detergent/protein ratio >500 (table 2). Additionally a group of additives commonly known to influence protein refolding, such as L-arginine, glycerol and guanidinium hydrochloride (all at various concentrations), were included in the refolding screen, but had no beneficial effect on refolding yields (data not shown). A number of different ionic milieus were also tested through the addition of ammonium acetate or potassium chloride, but also failed to increase the refolding efficiency (data not shown).

While the screen yielded acceptable refolding efficiencies for Y1L1 and Y1L2, Y1L3 and Y1L4 could only be refolded to very minor extents under the screened conditions. We therefore tried to use the most promising conditions as a starting point for further improvements. Because we had already tested a large number of different refolding buffer conditions our efforts were directed at improving refolding yields of promising starting conditions by using different refolding techniques for a given buffer composition. Initially our refolding screens were conducted by rapidly diluting the denatured protein in 8 M urea into the refolding buffer. We first wanted to know, whether urea as a denaturant might be responsible for the low refolding efficiency. To test this hypothesis the Y1L3 and Y1L4 constructs were re-expressed and purified this time in guanidinium hydrochloride (GdnHCl) and not urea as denaturant. It was found, that refolding from the GdnHCl state didn't have any influence on the refolding efficiency (data not shown).

As a next parameter the mixing speed of the denatured proteins with the detergent buffer was varied. Refolding was carried out by mixing the urea or GdnHCl denatured samples as slowly as possible by means of a peristaltic pump with the refolding buffer. This strategy resulted in significantly increased refolding yields for the Y1L3 and Y1L4 constructs (figure 12), suggesting that the presence of high amounts of unfolded proteins at the same time might interfere with their correct folding. Additionally lowering the temperature of the refolding reaction to 4 °C showed some beneficial effects.

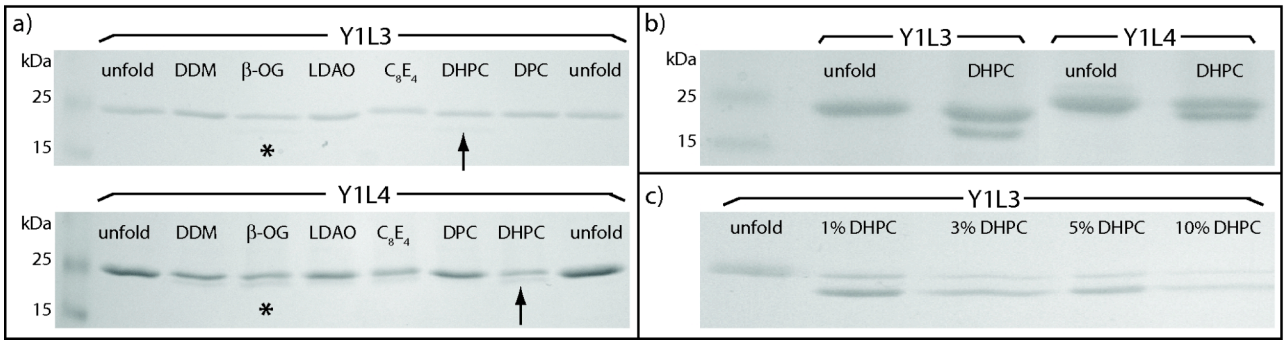


Figure 12: *Refolding optimization for the Y1L3 and Y1L4 constructs. "Unfold" refers to the heat-denatured form of the respective construct. a) Initial refolding trials for the two constructs at pH 10 in a number of different detergents. In DHPC (arrows) and β-OG (asterisk) a faint band of folded protein can be observed. b) Increased refolding yields after optimization of the refolding conditions and procedures in DHPC micelles. c) Refolding screen of Y1L3 in the presence of different DHPC concentrations.*

After this optimization Y1L3 and Y1L4 could be refolded to ~50%, which was enough to obtain samples of satisfactory quality for initial NMR studies. In the case of Y1L1 and Y1L2 these new refolding conditions didn't result in better refolding results, but rather gave less refolded product. Figure 13 depicts [^{15}N,^{1}H]-HSQC spectra of all four constructs: Y1L3 and Y1L4 were refolded unter the new improved conditions, while Y1L1 and Y1L2 were refolded according to the old protocol by simple rapid dilution into detergent containing buffer.

It should be noted here, that all four three-loop graft constructs showed a higher electrophoretic mobility in the folded state than in the unfolded state. This is opposite to what is observed for wt-OmpA, the one-loop exchange and the three-loop graft constructs. The presence of a larger number of sharp peaks in the random coil region in the NMR spectra of either the three- (figure 13) or one-loop graft constructs (figure 10) compared to wt-OmpA (figure 8) indicates that the wt-OmpA extracellular loops are less flexible than the Y1-loops present in the constructs.

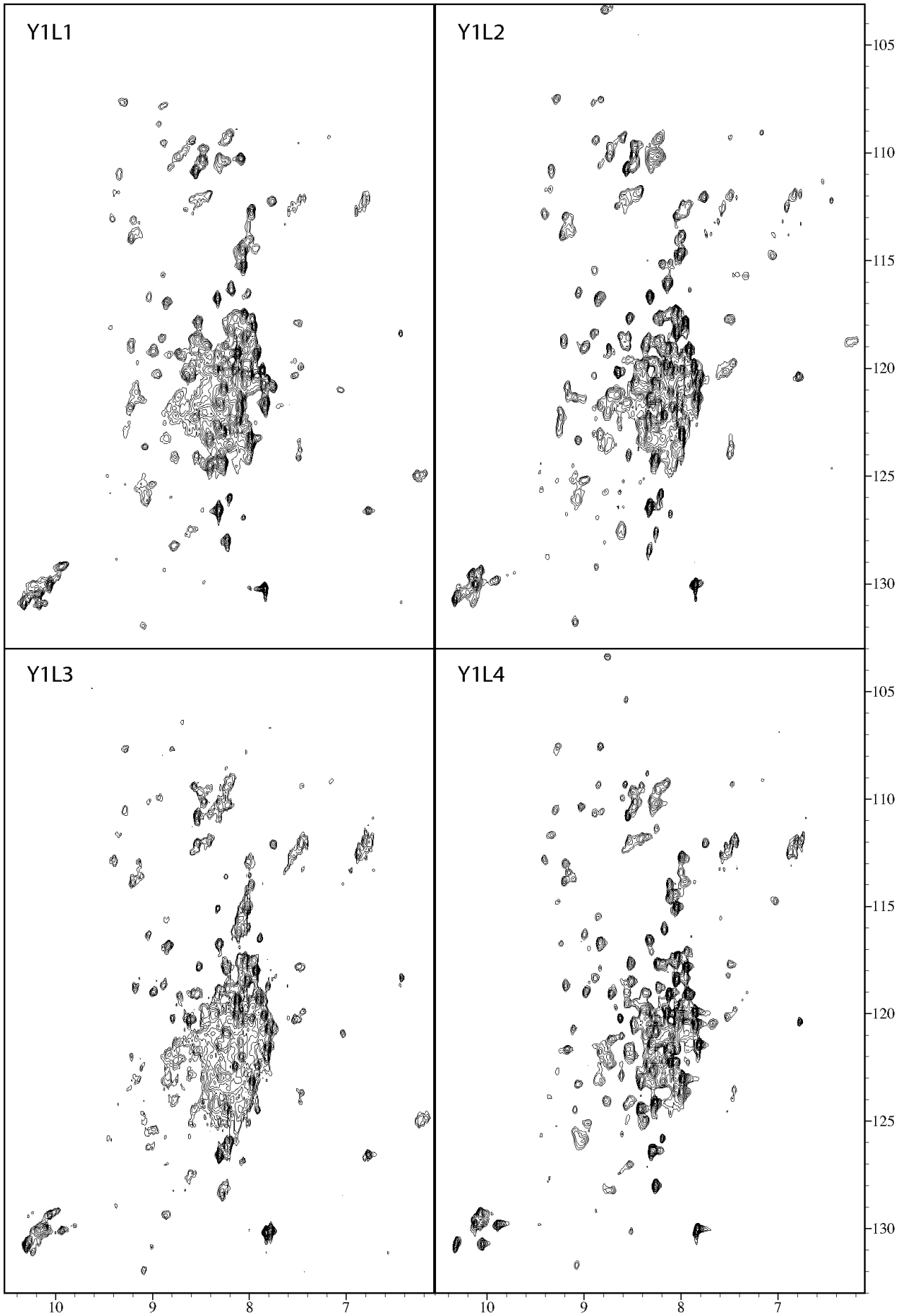


Figure 13: *[^{15}N,^{1}H]-HSQC spectra of the four three-loop graft constructs. Protein concentration was ~0.5 mM in 3% DHPC, 20 mM sodium phosphate, 100 mM NaCl at pH 6.5 and 47 °C.*

The high flexibility of the grafted loops results in a number of peaks in the NMR spectra, which are much sharper than expected for a protein-micelle complex of ca. 100 kDa. A similar behavior of the extracellular loops is also observed in wt-OmpA[8], however less pronounced than in the three-loop graft constructs (see figures 8 [OmpA] and 13 [three-loop graft constructs]). Assuming a sharp transition from a rigid β-barrel core, yielding broad resonance lines, to a much more flexible loop region, giving rise to sharp lines, one would expect ~100 broad signals from the β-barrel core and ~60 sharp signals from the loops for the three-loops graft constructs. The actual number of observed sharp peaks is ~40 and for the broad ones ~100 in the case of Y1L3. The non-observable resonances are probably rendered invisible by dynamics on the milli- to microsecond timescale, broadening their lines beyond detection limits.

Analysis of the refolding behavior by SDS-PAGE indicate an alteration in the folding efficiency upon exchanging native OmpA loops with the loops of the Y1-receptor. The folding mechanism of OmpA has been studied in great detail and the extracellular loops are thought to play a significant role in this folding process[25,76,77]. It is therefore not surprising that changes in those loops have an influence on the folding behavior. Preliminary NMR studies corroborate these findings. Spectra of the OmpA mutants are generally of lower quality than the ones of wildtype OmpA.

2.3 Interaction studies

The prime goals of our grafting attempts were: i) being able to obtain a stably folded chimeric receptor molecule displaying all three extracellular loops of a Y-receptor and ii) obtaining a system in which ligand binding affinity of the grafted loops can be observed, even when outside their natural 7 TM context. The accomplishment of the first aim could be demonstrated for the four three-loop graft constructs based on the analysis of the electrophoretic mobility and the signal dispersion in NMR experiments. In order to test whether our constructs exhibited ligand binding capability we chose to carry out chemical shift mapping experiments. For that purpose for at least one of either the ligand or the receptor constructs resonance assignments should be available.

2.3.1 Assignment of the neurohormones bound to DHPC micelles

For performing preliminary interaction studies, we decided to assign the backbone amide resonances of the peptide ligands of the Y-receptors in the presence of DHPC micelles, because these 36 amino acid peptides can be assigned relatively easily. Assignments of the neurohormones bound to DPC micelles are available, but only for very few resonances could this assignment be transferred to the spectra recorded in the presence of DHPC micelles. Briefly, spin system types were identified using a ^{15}N-resolved TOCSY experiment. Based on a NOE-relayed HSQC

experiment, these spin systems were linked into fragments, which in turn could be unambiguously assigned once they consisted of two or more spin systems. Complete assignment of the backbone amide resonances could be obtained for pNPY, two mutants of pNPY (pNPY-R33L and pNPY-R35L), pPYY and bPP bound to DHPC micelles. A table of all the assigned the chemical shifts of the neurohormones can be found in appendix H.

2.3.2 Chemical shift mapping studies with ^{15}N-labeled neurohormones and unlabeled three-loop graft constructs

With the assignment of the peptide ligands in hand we conducted preliminary interaction studies between the ligands and our receptor constructs. To this end we titrated ^{15}N-labeled neurohormones with unlabeled receptor constructs and observed the changes in the [^{15}N,^{1}H]-HSQC spectrum of the ligand. In figure 14 the changes in the spectra of NPY upon addition of 2 and 20 equivalents of the receptor constructs are shown. Whereas for the addition of Y1L1 and Y1L2 the spectra of NPY show no changes, Y1L3 and Y1L4, when added in excess, induce profound alterations to the spectra of NPY.

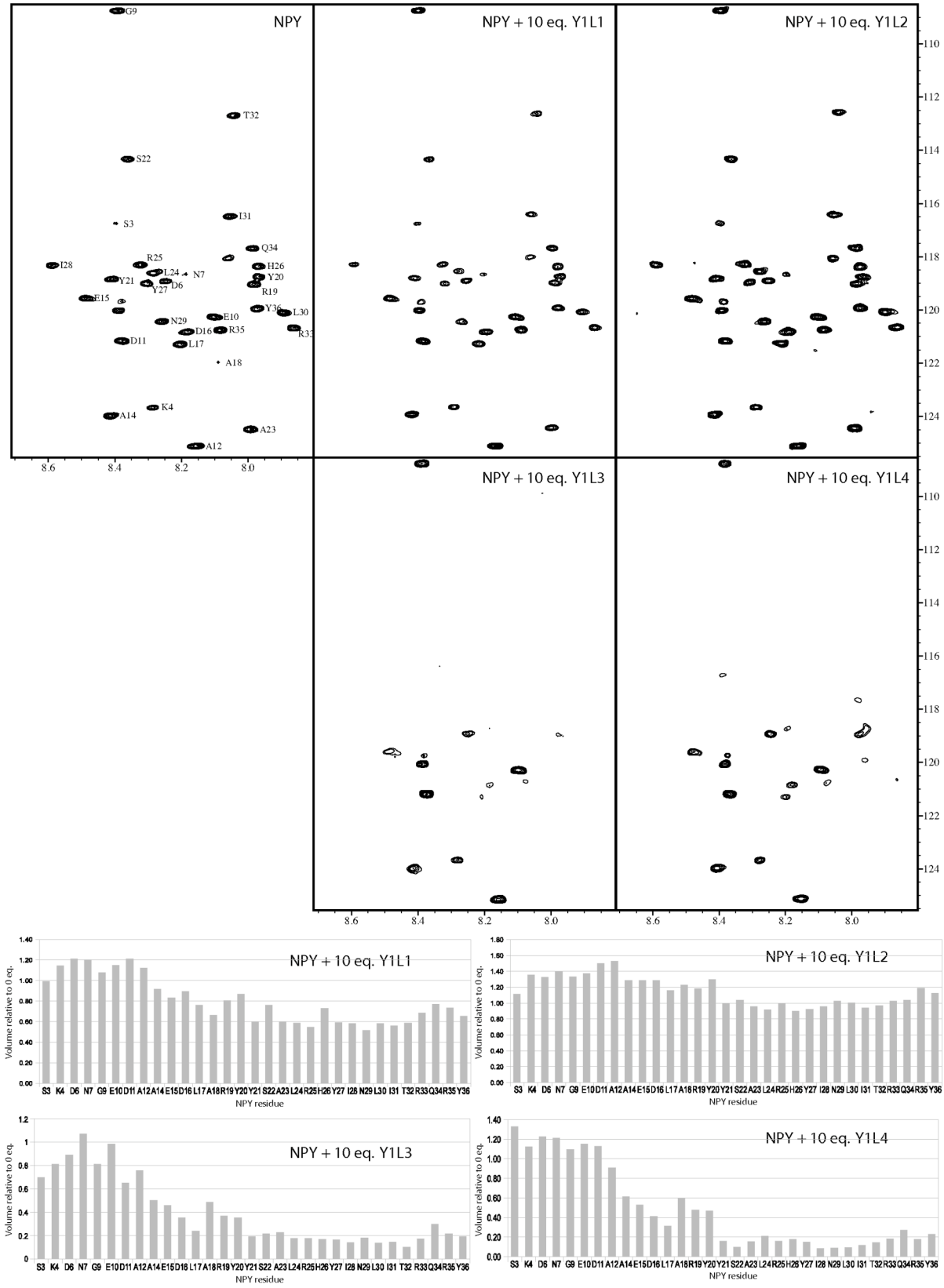


Figure 14: *[^{15}N,^{1}H]-HSQC spectra of NPY and of NPY in the presence of 10 equivalents of the four three-loop graft constructs. A clear difference in the spectra is seen after addition of Y1L3 and Y1L4. Below the spectra the peak volumes of the NPY resonances in presence of 10 equivalents relative to NPY without the receptor constructs are plotted. Whereas Y1L1 and Y1L2 show only small changes, the volumes of the C-terminal residues of Y1L3 and Y1L4 are reduced to below 20% of their initial values.*

In our case no shift in peak position but a decreased in peak intensity was observed, which is only consistent with a **slow** exchange process: A ligand that is in equilibrium between a bound and free state is expected to show different chemical shifts in the two states. If the exchange rate between the bound and free form is much smaller than the difference in chemical shift between the two states, two separate resonances per individual spin are expected: one arising from the free and one from the bound state (slow exchange). If, however, the exchange rate is much bigger than the difference in chemical shift between the two states, only one resonance can be observed (fast exchange regime). The position of that resonance will be shifted depending on the equilibrium constant for the exchange process. In intermediate cases, where the rate constant of the exchange process is in the range of the chemical shift difference between the two states, a pronounced increase in the line widths of the resonances can be observed. This "exchange broadening" can broaden a peak beyond detection. The excessive broadening of a resonance leading to its disappearance from the spectrum could be explained either through such an unfavorable exchange or a dramatic increase in apparent molecular weight of the ligand, when it is in complex with a receptor construct.

Because of the fact that no shift in peak position occurs and because the remaining peak is not particularly broadened, we propose that the remaining peak intensity is due to the unbound ligand. The signal due to the bound ligand is shifted, as expected in slow exchange, but broadened due to intermediate exchange beyond detection. In the alternative case of fast exchange a shift in position must occur.

Fundamentally the line width of an NMR signal is dependent on the tumbling rate (quantified as the correlation time, τ_c) of the molecule from which it arises. In a solvent with a given viscosity the tumbling rate in turn is primarily determined by the size of the molecule. In principle, complex formation may increase the size of the mixed micelle so much that additional, strong line-broadening could occur. However, the fact that resonances of the protein are observable argues against this, and only leaves the possibility that complex formation leads to recruitment of more than one loop-grafted protein to the complex, which is highly unlikely. Instead, the disappearance of the C-terminal peaks is more likely due to an exchange process that may be due to small conformational exchange processes in the receptor-bound state.

Integration of the peaks in the [^{15}N,^{1}H]-HSQC allows to quantify the changes in the ligand's signals upon addition of receptor construct. In figure 14 the volume changes of the NPY signals upon addition of an excess of receptor constructs are shown. The plots show clearly that the C-terminal residues are much more affected than the N-terminal ones. This implies a binding of the C-terminal α-helix of NPY to the receptor construct, while the N-terminus is moving freely in solution. This is in agreement with the identification of the C-terminal residues of NPY being more important for receptor binding than the N-terminal ones[35,58]. Titration of the other two members of the NPY

family of neurohormones, PP and PYY, gave similar results (figure 15). Whereas NPY shows a very clear difference in the effects of the Y1L3 and Y1L4 receptor constructs on the N- and C-terminus, this behavior is less pronounced with PYY and particularly with PP. Considering that the binding profile of PP to the Y receptor subtype is different than from NPY and PYY, this may in general indicate a slightly different binding mode.

Figure 15: *Interaction profiles of pNPY (blue), pPYY (orange) and bPP (green) with a 10-fold excess of Y1L3.*

In our eyes the specificity of the interaction between the Y1L3/Y1L4 receptor construct pair and the neurohormones of the NPY family is corroborated by the fact, that both Y1L1 and Y1L2 don't show a significant interaction with the neurohormones, and can therefore by considered as negative controls. Also the wt-OmpA and the minimal length OmpA did not have any influence on the spectra of the neurohormones (data not shown). To gain additional support for this view we conducted a competition experiment, in which unlabeled NPY competes for the binding to unlabeled receptor construct with ^{15}N-labeled NPY (figure 16). To a sample of ^{15}N-NPY in the presence of 10 equivalents of unlabeled Y1L3 an excess of 50 equivalents of unlabeled NPY was added. The re-appearing of all the NPY resonances in the [^{15}N,^{1}H]-HSQC shows that the unlabeled NPY competes with ^{15}N-NPY for the binding to the Y1L3 receptor construct. As negative controls the effect of Y1L3 on the spectra of ubiquitin and two NPY mutants (NPY-R33L and NPY-R35L) with reduced binding affinity to the Y-receptor *in vivo*[35] were investigated. The spectra of ubiquitin showed no change upon addition of several equivalents of Y1L3 (data not shown). The two binding deficient NPY mutants showed a markedly decreased affinity for Y1L3 (figure 16). In summary these findings confirm our view of a specific interaction of the neurohormones of the NPY family with the receptor constructs Y1L3 and Y1L4.

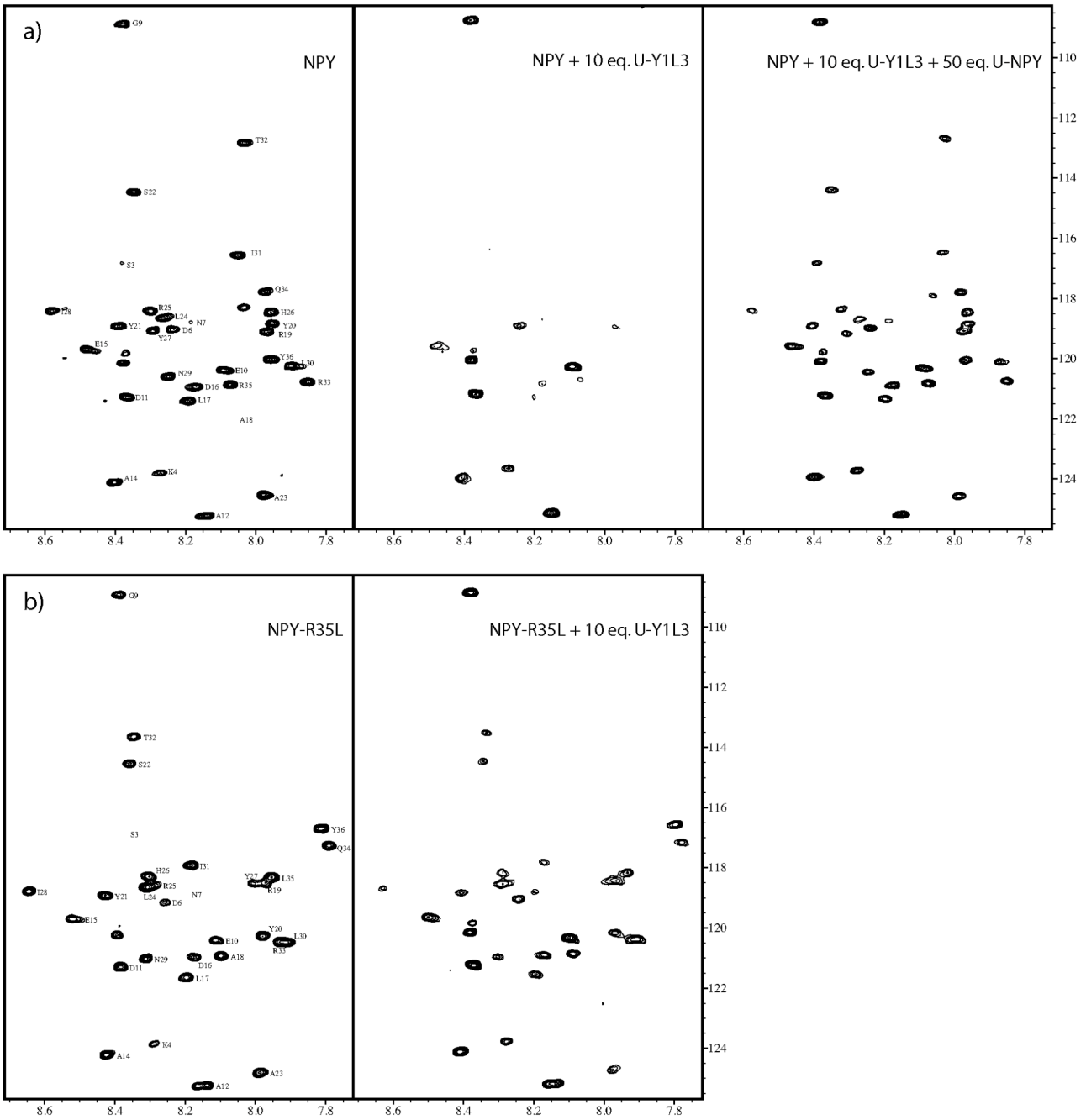


Figure 16: *Specificity of the interaction of NPY with Y1L3. a) Unlabeled NPY is able to compete with ^{15}N-labeled NPY for the binding to Y1L3. Addition of 50 eq. of unlabeled NPY (right spectrum) to a sample of ^{15}N-NPY (left spectrum) treated with 10 eq. of unlabeled Y1L3 (middle spectrum) leads to reappearance of the resonances from ^{15}N-NPY. b) A mutant of NPY (NPY-R35L) with a known decreased binding affinity to the Y1-receptor in vivo shows weaker interaction with Y1L3 than does wt-NPY.*

Because the refolding efficiency of the four constructs is less than 100% it is not *a priori* clear that it is the folded component of the mixture, which is interacting with the neurohormones. Given the lack of a method to separate the folded from the unfolded construct and the failing attempts at trying to achieve 100% refolding, we chose to "refold" Y1L3 under unfavorable conditions (resulting in almost 100% unfolded Y1L3 as judged by SDS-PAGE) and testing the effect of this unfolded form of the receptor construct on the spectra of NPY. We can proudly say that from the

screening of refolding conditions we had a large number of such unfavorable conditions at hand. The most appealing seemed to carry out the refolding in the same detergent that yielded good refolding results, but instead of the favorable pH 10 at the unfavorable pH 4. However, at pH 4 the receptor constructs mainly precipitated from the detergent solution, leaving no micelle integrated/associated misfolded protein present in solution. We therefore chose to carry out the "refolding" at the favorable pH 10, but using an unfavorable detergent. In the zwitterionic detergent DPC the refolding efficiency of Y1L3 was essentially zero also at pH 10, without showing any protein aggregates precipitating out of the solution. NPY was titrated with Y1L3 refolded in DPC at pH 10 and addition of 15 equivalents of unfolded Y1L3 to NPY had much less effect than the addition of 10 equivalents of the partially folded Y1L3 (figure 17). This lead us to believe that it should be indeed the folded fraction of the receptor constructs, which is responsible for the interaction with the neurohormones.

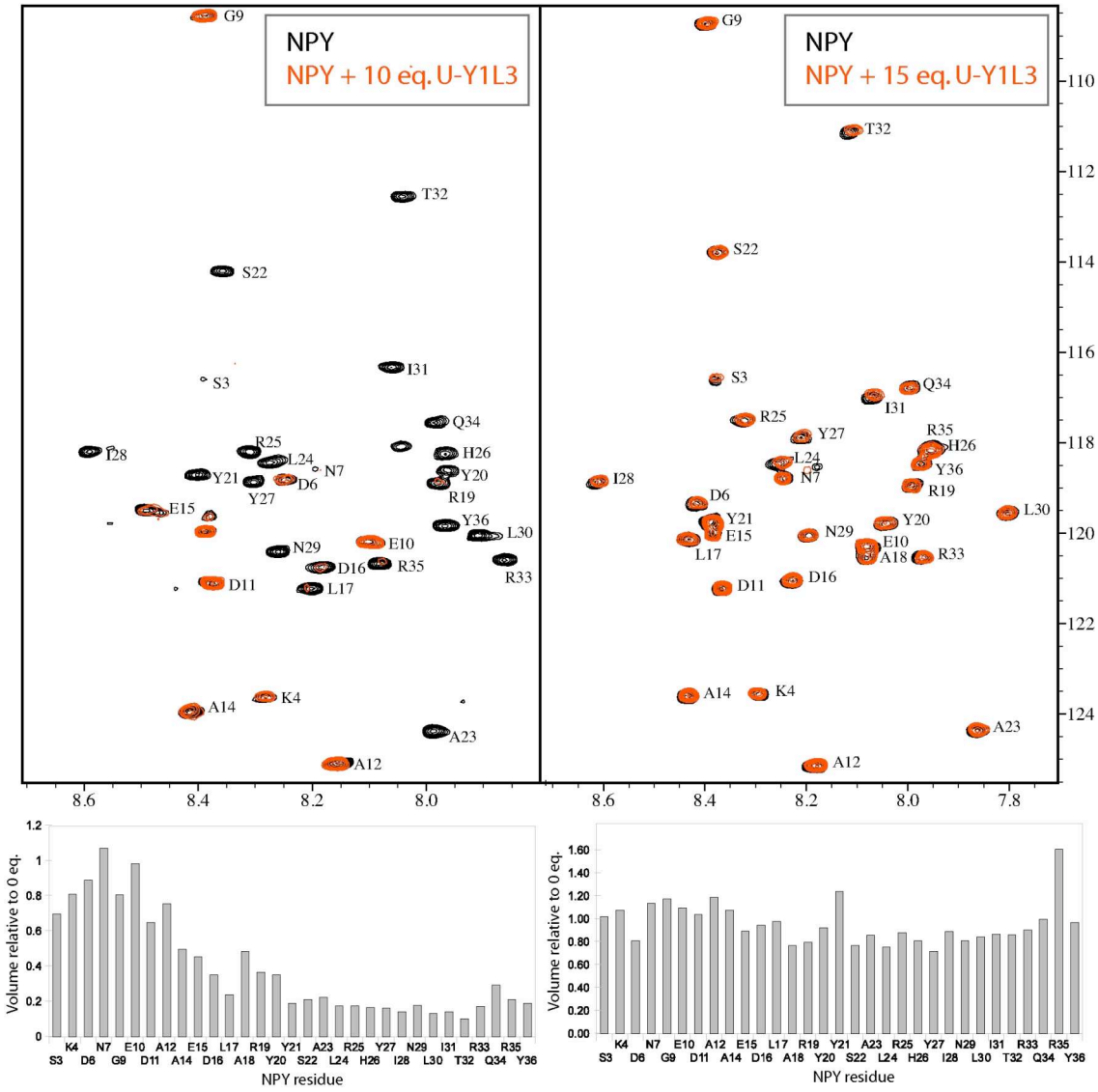


Figure 17: *Comparison of the interaction profile of NPY with a mixture of folded and unfolded Y1L3 (in DHPC; left) and completely unfolded Y1L3 (in DPC; right).*

2.3.3 Chemical shift mapping studies with ^{15}N-labeled three-loop graft constructs and unlabeled neurohormones

To obtain complementary information about the interaction we sought to titrate the ^{15}N-labeled receptor constructs with unlabeled neurohormones and observe, what changes are induced in the spectra of the receptor. To this end we have expressed all four receptor constructs in ^{15}N-labeled form. They were refolded into DHPC detergent micelles and titrated with unlabeled neurohormones (figure 18). Surprisingly the [^{15}N,^{1}H]-HSQC spectra of the receptor constructs do not show any changes in the peak patterns upon addition of NPY. A possible explanation for this unexpected behavior is, that the residues of the receptor constructs, which are responsible for the interaction with the neurohormones, are not visible in the spectra. That this might be indeed the case is corroborated by the observation that the actual number of observed sharp (and thus most likely originating from the flexible loops) peaks in the receptor construct is ~40 out of an expected 59. The non-observable peaks are probably absent due to dynamics on the milli- to microsecond timescale, broadening their lines beyond detection limits.

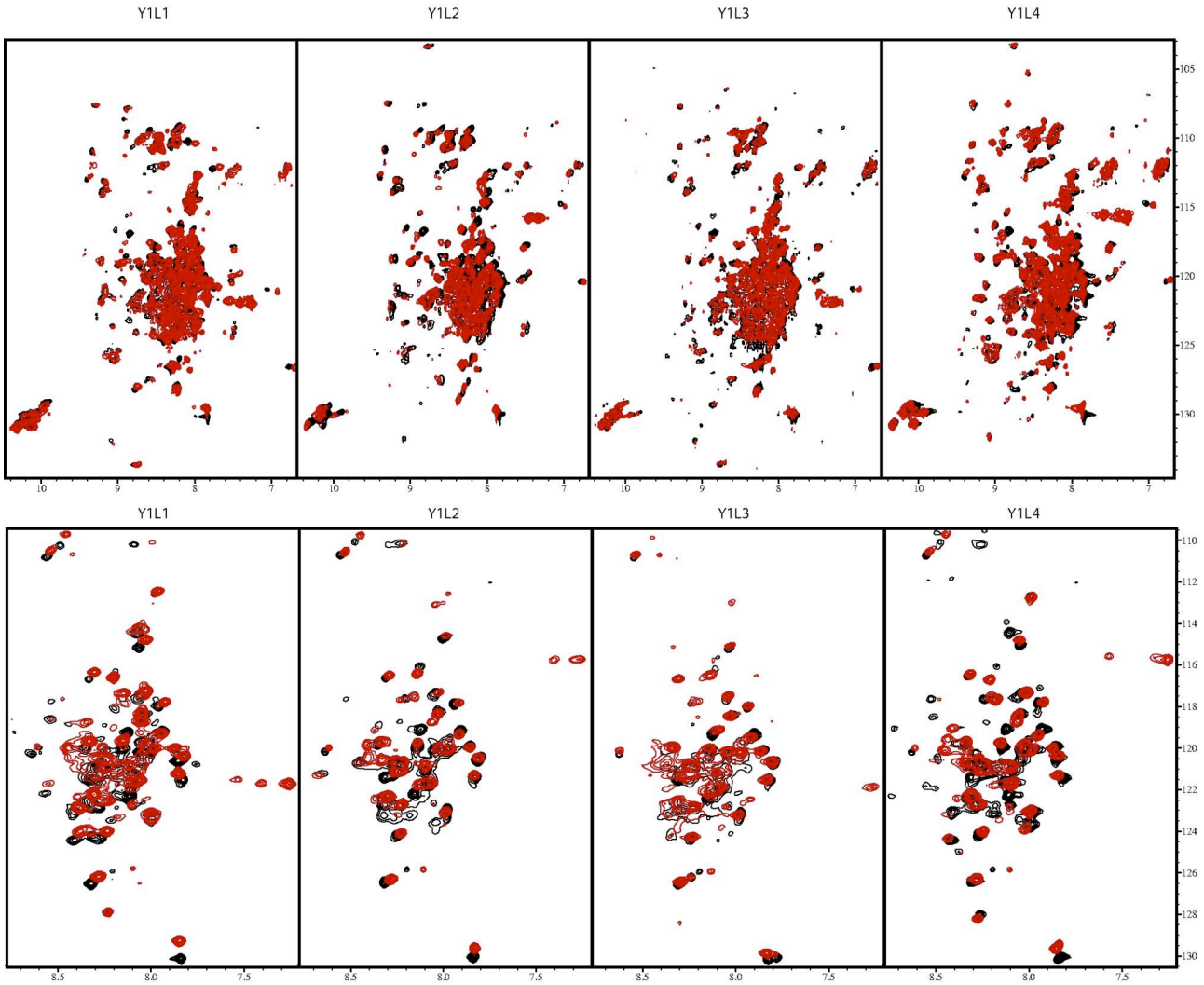


Figure 18: *Interaction studies of ^{15}N-labeled receptor constructs with unlabeled NPY. HSQC spectra covering the full amide proton chemical shift range (upper row) and zoomed in the random coil region of the amide proton chemical shift range (lower row) are depicted in the absence (black) and in the presence (red spectra) of 10 equivalents of unlabeled NPY.*

2.3.4 Backbone assignment of Y1L3

In order to get to know, which residues are *not* involved in the binding of the neurohormones, we set out to obtain resonance assignments for as many signals in the Y1L3 receptor construct as possible. The commonly used assignment strategies for membrane proteins of this size are based on triple resonance experiments and require a high level of deuteration. To this end, we expressed Y1L3 in triple labeled minimal medium containing $^{15}NH_4Cl$ and deuterated glucose ($^{2}H_7$,$^{13}C_6$-D-Glucose) as the sole nitrogen and carbon source, respectively. 99% D_2O served as the solvent. With this expression strategy a high level of deuteration can be achieved. In order to allow proton NMR experiments ^{1}H has to be re-introduced at the exchangeable amide nitrogens. In our case this is achieved by purifying the protein under denaturing conditions in H_2O-based solvents. The high level of deuteration is absolutely necessary for triple resonance experiments with membrane

proteins of this size. The comparison between a fully protonated and a highly deuterated sample shown in figure 19b illustrates the dramatic improvement in signal-to-noise ratio brought about by high levels of deuteration.

Backbone assignment was carried out with help of a HN(CO)CACB and a HNCACB experiment. Out of the 32 expected backbone amide resonances of the long e2-loop (34 amino acids, 2 of which are prolines) 31 could be assigned. Of the shorter e1- and e3-loops (13 and 14 expected resonances, respectively) only 3 backbone amide resonances from each loop could be assigned. This shows that the exchange broadening is exclusively a problem for these two loop sequences. We think we can exclude the possibility, that the absence of those loop resonances is a molecular weight effect, because we can see many of the peaks of the β-barrel in our spectra.

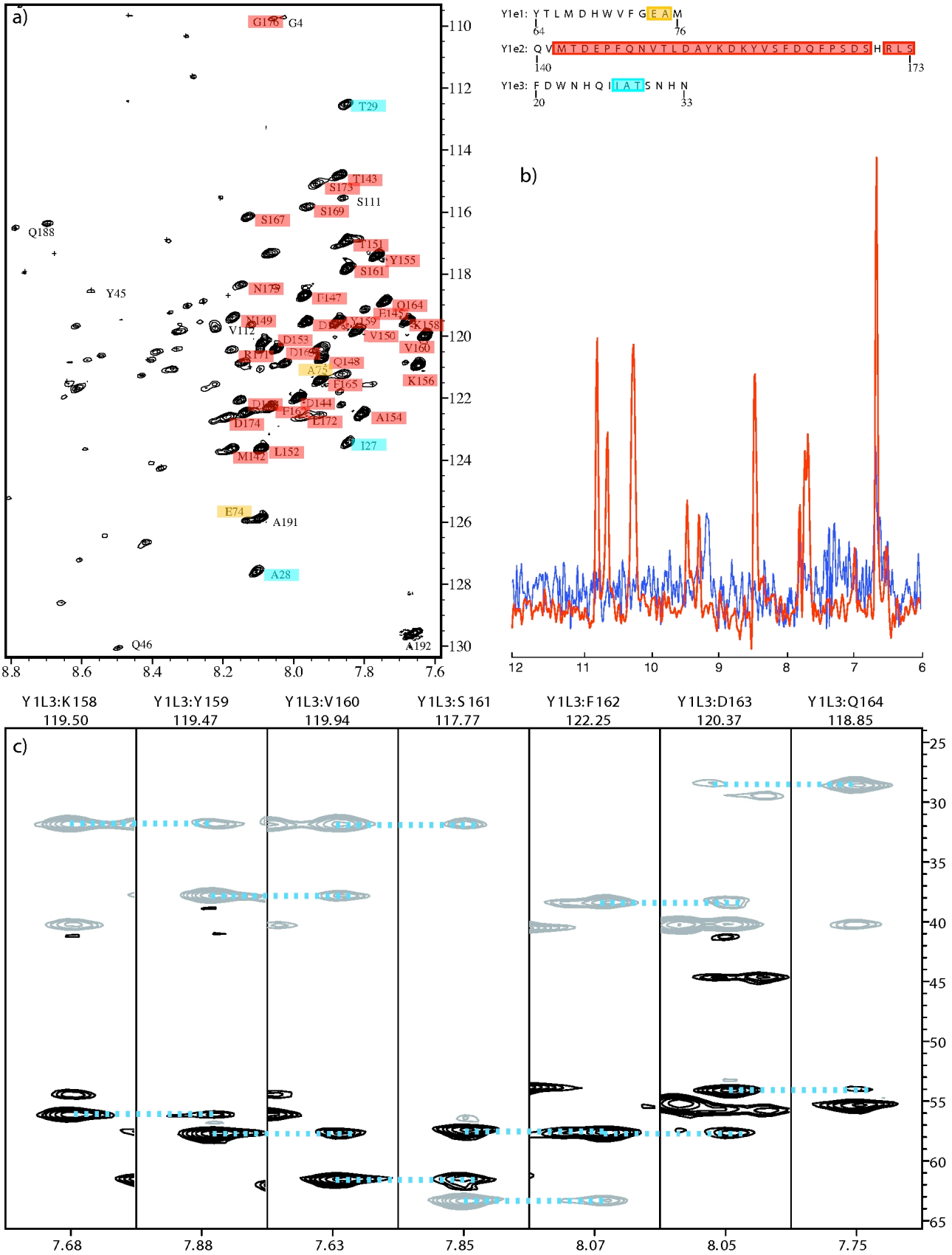


Figure 19: *Assignment of Y1L3. a) Assignments of grafted Y1-receptor loops are highlighted in yellow (e1-loop), red (e2-loop) and blue (e3-loop). b) The effect of deuteration on the signal-to-noise ratio illustrated by a slice through the [^{15}N,^{1}H]-HTROSY spectrum at a ^{15}N chemical shift of 112.5 ppm of a protonated (blue) and deuterated (red) sample. c) Assignment strategy based on triple resonance spectra illustrated on a series of matching HNCACB strips.*

2.3.5 The effect of increasing the flexibility of the grafted loops

We tried to increase the flexibility of the grafted loops - primarily of the e1- and e3-loop sequences - by inserting flexible glycine-serine linkers of different length in between the anchor sites on the scaffold and the Y1-receptor's extracellular loops. Derived from the construct Y1L3 we generated 4 derivatives: one in which each Y1-receptor loop was flanked at both ends by a Ser-Gly dipeptide and one in which it was flanked by a Ser-Gly-Ser-Gly tetrapeptide (Y1L3-GS and Y1L3-GSGS). Furthermore two analogous constructs, in which only the e1- and e3-loops were flanked, but not the e2-loop, were created (Y1L3-gs and Y1L3-gsgs). We observed a significantly different refolding behavior of the different linker-flanked constructs. Only the two constructs carrying the flanking linker sequences at all three loops were able to fold equally well as the parent Y1L3 construct. Whereas the refolding efficiency of these constructs was similar to Y1L3 (~80% efficiency) the two constructs carrying the linker sequences only on the e1- and e3-loop only folded with an efficiency of only ~50%. For the sake of completeness we also expressed constructs carrying the di- or tetrapeptide linker sequences only at the e2-loop (Y1L3-e2gs and Y1L3-e2gsgs). These showed the same folding deficiency that was observed for Y1L3-gs and Y1L3-gsgs (figure 20).

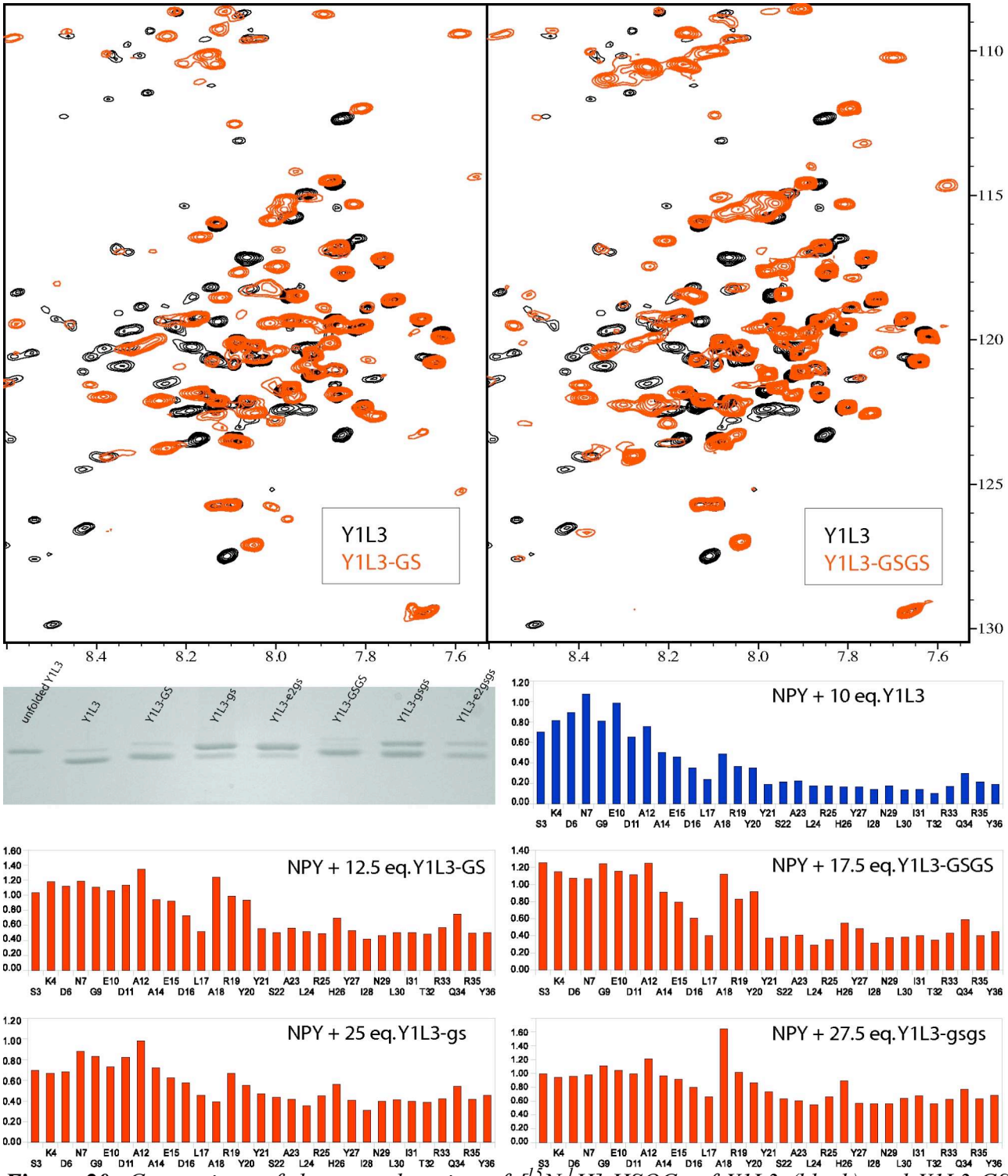


Figure 20: *Comparison of the central region of [^{15}N,^{1}H]-HSQCs of Y1L3 (black) and Y1L3-GS (red, left side) and Y1L3-GSGS (red, right side). A comparison of the refolding behavior and the interaction profiles of the four constructs Y1L3-gs, Y1L3-gsgs, Y1L3-GS and Y1L3-GSGS with that of Y1L3 is shown below the NMR spectra.*

The [^{15}N,^{1}H]-HSQC spectra of Y1L3-GS and Y1L3-GSGS show additional peaks relative to Y1L3. Most new peaks, however, are in the region of the spectrum, which is characteristic for the glycine and serine resonances, and the peaks thus most likely arise from the glycine-serine linkers themselves, rather than from the intervening loop sequences. Overlays of the spectra from Y1L3

and both Y1L3-GS and -GSGS are shown in figure 20. Not surprisingly the glycine-serine linker flanked constructs showed a decreased affinity towards the neurohormones (figure 20).

2.3.6 Ion exchange purification of 100% folded three-loop graft constructs and chemical shift mapping studies with neurohormones

We would like to repeat here, that all the interaction studies carried out between the neurohormones and the receptor constructs have been performed with a mixture of folded and unfolded receptor constructs. Even though we had indications, that it is indeed the folded fraction of the construct which is responsible for the interaction with the neurohormones, a definitive proof would only be the repetition of the experiments with 100% folded receptor constructs. It is possible to resolve the unfolded from the folded species on an analytical scale by means of non-denaturing SDS-PAGE. It should be possible therefore to achieve the same thing, on a preparative scale, by means of size exclusion chromatography. The application of this technique, however, requires large column volumes. Especially if the effective size difference that should be resolved, is small. This seems to be the case for the receptor constructs, as a resolution of the two folding states by SDS-PAGE was only possible after careful optimization of the conditions, under which the SDS-PAGE was performed. Since we are working here with a membrane protein, the presence of a detergent above its critical micellar concentration at all times is critical to the stability of the folded protein. The high cost of DHPC, in which the receptor constructs showed the highest fraction of folded protein, precludes the use of this detergent in size exclusion chromatography.

Ion exchange chromatography is a method for separation of molecules based on their charge. This seems to exclude this technique for the separation of two isoforms of the same protein, as both have exactly the same primary sequence, and hence have identical numbers of charged residues. However, differently folded protein species are expected to display on their surface different numbers of charged groups, which are moreover in different orientation relative to each other, resulting in distinct interaction strengths with the charged groups of an ion-exchange resin. The small column volumes used in ion exchange chromatography makes this technique also appealing to detergent containing buffers. With isoelectric points at pH 5.5 and 4.7 for wt-OmpA and the receptor constructs, respectively, these molecules are charged negatively in all our refolding and further processing steps. To make use of ion exchange chromatography for the separation of negatively charged molecules anion exchange chromatography is the method of choice. Our refolding has always been carried out in borate buffer at pH 10. As borate is charged highly negatively at this pH it is not a good buffer salt for anion-exchange chromatography. We therefore switched the refolding buffer to piperazine or glycine, both also at pH 10.

Initial trials were carried out with wt-OmpA and the cheap detergent LDAO in which wt-OmpA

folds to about 50% at pH 10 (figure 8). This mixture was loaded onto a monoQ anion exchange column with a column volume of 1 ml, and bound molecules were eluted with a NaCl gradient. The chromatogram of wt-OmpA in LDAO does not show two clearly resolved peaks. Analysis of the collected fractions, however, showed that folded wt-OmpA could – at least partially – be separated from unfolded wt-OmpA (data not shown). With this result we set out to try our luck with the receptor construct Y1L3. The refolding efficiency of this construct in LDAO micelles is only around 10% as judged by the relative band intensities on polyacrylamide gels. However, this should be sufficient for the assessment of the usability of ion exchange chromatography in separating folded from unfolded Y1L3. The chromatogram of this purification is shown in figure 21a along with the SDS-PAGE analysis of the collected fractions. In the chromatogram at around 125 mM NaCl a sharp peak is seen in the UV absorption. SDS-PAGE revealed this peak to correspond to the folded Y1L3, while a broader peak at 400-500 mM NaCl represents the unfolded form(s) of the receptor construct.

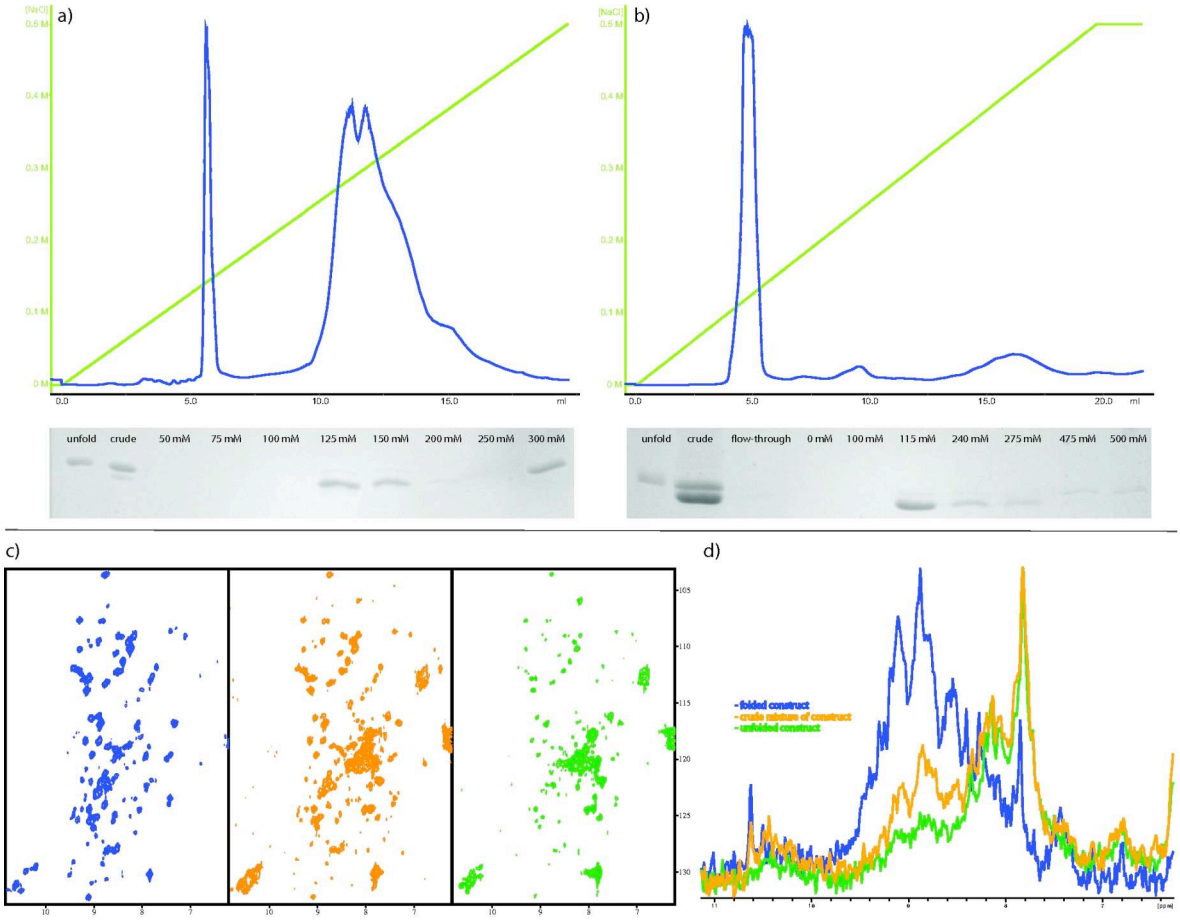


Figure 21: *a) Anion exchange chromatogram of Y1L3 in LDAO micelles. The sharp peak at ~125 mM NaCl corresponds to the folded form. b) Anion exchange chromatogram of Y1L3 in DHPC micelles. The sharp peak at ~115 mM NaCl corresponds to the folded form. c) Comparison of the spectra of the folded form of Y1L3 (blue) with a mixture of folded and unfolded (orange) and the unfolded form of Y1L3 (green). d) Same comparison as in c), but on the 1D projections of the spectra from c). The projection of the crude mixture is shown to be approximately the sum of the projections of the folded and unfolded forms.*

The resolution of the anion-exchange purification is thus much better for the Y1L3 construct than for the wt-OmpA. The same behavior was observed also for Y1L4 (data not shown). A plausible explanation for this different behavior between wt-OmpA and Y1L3/Y1L4 might be, that the receptor constructs display 5 negatively charged groups more and 1 positively charged group less in their extracellular loops than does the wt-OmpA. Encouraged by this promising results observed with the LDAO detergent, we attempted the same separation with DHPC as the detergent and obtained similar results (figure 21b). We where thus finally able to obtain the receptor constructs in 100% folded form. That the fractions eluting at ~125 mM NaCl contain indeed properly folded protein, is corroborated by comparing the chemical shift dispersion observed in these fractions compared to the unfolded fractions eluting at 400-500 mM NaCl. Figures 21c and 21d show [^{15}N,^{1}H]-HSQC spectra and 1D-projections of the folded and unfolded fractions in comparison with spectra from the crude mixture of folded and unfolded protein as is present before the chromatography. The spectrum of the crude mixture corresponds about to the sum of the spectra from the folded and unfolded forms.

The titrations of the neurohormones with this folded form, however, didn't show any sign of interaction (figure 22). Likewise, no interaction between the unfolded fractions and the neurohormones could be observed. Only when all fractions from the purification were mixed, the characteristic decrease in the peak volumes of the C-terminal residues of the neurohormones could be observed.

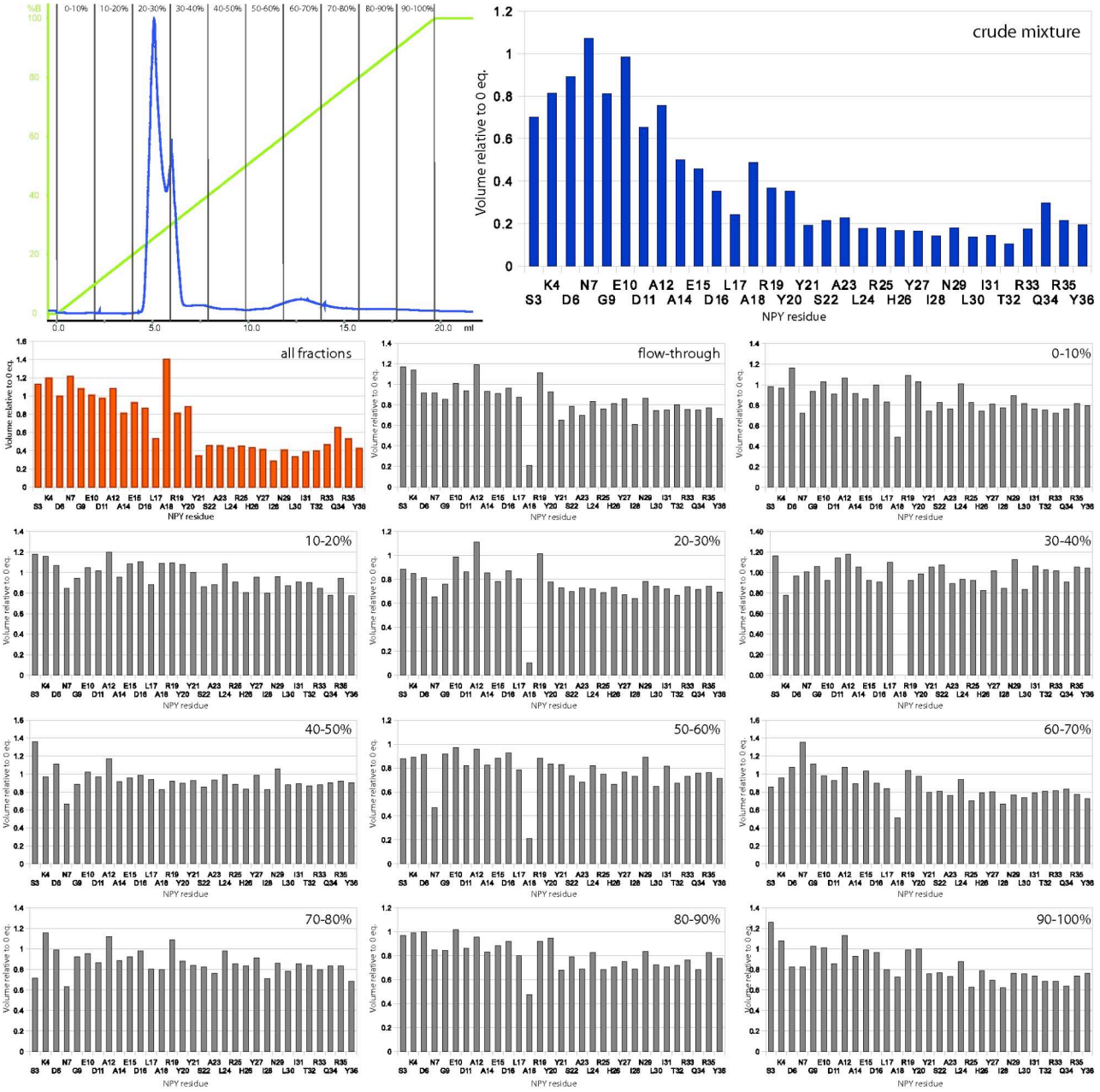


Figure 22: *Comparison of the interaction profiles of NPY with all the fractions collected during an ion-exchange purification run. Top left: An anion-exchange chromatogram of the purification of Y1L3 in DHPC. Bound protein was eluted with a gradient from 0 to 1 M NaCl and fractions were collected for all intervals as indicated by the vertical gray lines. Top right: Interaction profile of NPY with a crude mixture of folded and unfolded Y1L3. Bottom: Interaction profiles for each of the collected fractions indicated in the chromatogram, the flow-through and a mixture of all fractions. As can be seen none of the individual fractions (gray profiles) is capable of eliciting an interaction profile as observed with the crude mixture. Only when all fractions are reunited (orange profile), the interaction profile of the crude mixture can be reproduced.*

3. Discussion

3.1 Design of chimeric receptor constructs

In this work we describe the attempt at grafting the extracellular loops of a GPCR onto the TM β-barrel scaffold of a minimal length OmpA to generate a chimeric receptor. The transfer of a given functionality from one protein to another is a challenging task. It is most likely only promising under the assumption, i) that all the essential residues responsible for that functionality are transferred and, ii) that these residues retain a suitable relative orientation. In order to obtain a chimeric receptor, displaying (natural) ligand binding ability, the distances between loop anchor points and between individual loops found in the receptor should be matched by the scaffold as closely as possible. The GPCRs of our interest are the receptors of the NPY family of neurohormones, the so-called Y-receptors[78]. Unfortunately to date only little structural information on GPCRs is available and none of the Y-receptors is among those GPCRs for which a crystal structure could be obtained. In our study we thus relied on the assumption of structural homology between the known GPCR structures and the Y-receptors. This hypothesis is common in the field of molecular biology and has formed the basis for many homology modeling studies of GPCRs[79].

In a selected set of GPCR structures, the spacing between the anchor points for the three extracellular loops is at an average of 13 Å for e1 and e2 and 14 Å for e3 and shows a narrow distribution of roughly +/- 3 Å. The distances between the anchor points not directly connected by an extracellular loop are less well conserved. This indicates, that the relative positions of two helices anchoring an extracellular loop are more conserved than the relative positions between helices not directly connected by an extracellular loop. The observed average values over the distances for the analyzed GPCRs served as an estimate for the distances which can be expected in the Y-receptors, for which up till today no high resolution structural data is available.

The β-barrel TM domain of OmpA has been solved both by X-ray crystallography[6,7] and NMR spectroscopy based on NOE restraints[8] and refined using residual dipolar couplings in addition to NOE restraints[9]. A study using the β-barrel core of OmpA as a scaffold for the Ca^{2+}-binding EF-hand loop from *X. laevis* has yielded a high resolution NMR structure as well[28]. Not surprisingly all five of these structures show higher flexibility of the extracellular loop as compared to the β-strands of the barrel and the short periplasmic loops. It is not clear, whether this high mobility is also observed in nature or is - at least partly - an effect of the detergents used in the crystallographic (C_8E_4) and NMR (DPC or DHPC) studies, which are chemically very different from the lipopolysaccharide environment, which the extracellular loops are facing *in vivo*[6].

It seems clear, that any scaffold molecule will never perfectly match the loop geometries found in GPCRs. However, the a statistical analysis of the distances between the anchor points of the

extracellular loops in the NMR structures published on OmpA (pdb codes 1G90, 2GE4 and 2JMM) shows, that these distances are definitely within the same range as the ones also observed in the GPCRs. The relatively broad distibution of distances between the anchor point whithin one set of conformers of an NMR structure of OmpA indicates, that the anchor points of the β-barrel possess significant ability for structural adaptation. This indicates that the β-barrel of OmpA might indeed be a good scaffold for the extracellular loops of GPCRs. The possibility of inserting flexible linker residues between the scaffold's anchor points and the grafted loop residues of the Y-receptors presents an additional means of overcoming distance mismatches between the anchor points of the GPCRs and the ones of the scaffold.

Inspection of the crystal structures which have been published on GPCRs so far[63,80-85] shows, that the extracellular loops are quite flexible. They can hence be assumed to sample a large conformational space. The distance distribution between the anchor points of our selected scaffold seem to allow for such a flexibility as well.

We have chosen the Y1-receptor as the "donor" for the extracellular loops over the other Y-receptors, because the most biochemical data is published on that receptor subtype. As no crystal structure of the Y-receptors is available the identification of the alleged loop residues is based on structure prediction, GPCR sequence homologies and biochemical data. Based on these data the Y1-receptor loop sequences to be grafted were selected. Since OmpA contains 4 extracellular loops and GPCRs only 3, one acceptor position in the three-loop graft constructs had to remain empty. In this position a very short β-turn inducting sequence of 1-2 residues was inserted, in order to minimize the possibility of interference with the GPCR loops. A comparison of all peptide binding GPCRs shows, that Asp6.59 is the only residue which is only conserved in the extracellular loops in all subtypes of the Y-receptors (Y1 to Y5) and throughout all species (see GPCRs_all_peptide_ligand.alig). Interestingly in other peptide ligand GPCRs the positions 6.59 and 6.58 are often found to be also conserved within one family[67]. Position 6.59 lies at the top of TM6 and the beginning of ECL3. Studies on rhodopsin and other GPCRs showed that the top region of TM6 and the adjacent ECL are important in ligand binding[86-89].

The lenghts of the grafted loops were 15 residues for the e1 and e3 loops and 30 residues for the e2 loop. The lenght of the OmpA extracellular loops are also in this range (around 14-18 residues). It has been known that the formation of the β-barrel is compatible with very short sequences of only 1-2 residues[27] and with eukaryotic sequences of about equal length than the original OmpA extracellular loops[28]. It could therefore be expected that the scaffold would tolerate at least the e1 and e3 loop, which even though they are of eukaryotic origin, they have about equal length than the native OmpA loops. This is not the case for the much longer e2 loop. To test the compatibility of the extracellular looop sequences of the Y1-receptor with the β-barrel core of OmpA, we created

mutants in which each extracellular loop of OmpA was replaced with each of the Y1-receptor's loops. All 12 variants of these "one-loop exchange construct" could be expressed and refolded as the wt-OmpA, indicating the compatibility of the Y1-receptor loop sequences with our scaffold. In a second set of constructs one OmpA loop was replaced with a GPCR loop and the other three OmpA loops were replaced by a minimal β-turn inducing sequence. Also these "one-loop graft constructs" could be expressed and refolded, albeit with a lower efficiency than for the wt-OmpA and the one-loop exchange constructs. The partial loss of folding capacity thus seems to be an effect of the shortening of the loops and not an effect of the GPCR loops.

The arrangement of three donor looop sequences on the four acceptor sites of the scaffold is possible in 24 different ways. In order to rule out constructs with relative loop orientations deviating strongly from the average GPCR topology, we calculated a "mismatch score" for each of the 24 possible arrangements. Finally four of these "three-loop graft constructs", named Y1L1, Y1L2, Y1L3 and Y1L4, were assembled and expressed. Also they showed a reduced folding proficiency compared to OmpA. In addition, both the one- and three-loop graft constructs showed a different electrophoretic mobility behavior then the wt-OmpA or the one-loop exchange constructs. While for the latter two the folded species migrates slower compared to the unfolded species, this behavior is reversed in the former.

3.2 Folding studies of the receptor constructs

The presence of a larger number of sharp peaks in the random coil region in the NMR spectra of the graft constructs compared to wt-OmpA, indicates that the wt-OmpA extracellular loops are less flexible than the Y1-loops present in the chimeric constructs. The increased electrophoretic mobility of the three-loop graft constructs compared to the wt-OmpA might thus be explained by the higher flexibility of the Y1-receptor loops, imposing less sterical hindrance during the migration through a polyacrylamide gel. Based on the results of the refolding studies with the one-loop exchange contructs, this increased mobility is mainly mediated by the long Y1e2 loop, as this loop alone is capable of removing the the mobility difference between folded and unfolded forms (figure 9c). Other possible factors, such as differences in the overall length or charge of the grafted loops compared to the wildtype loops, cannot consistently explain the observed electrophoretic behavior.

Refolding screens were carried out for the three-loop graft constructs with a six different detergents, three different pH values and a variety of additivies, commonly known to influence folding efficiency. While none of the additives had great influence, the pH and the detergent was critical for successful refolding. This behavior is somewhat similar to what is observed for wt-OmpA, where also additives didn't show any effect and pH was critical. Also the detergent showed some influence on the folding of wt-OmpA, though less severe. A decreased refolding efficiency was also reported

in a previous study, which has used the β-barrel of OmpA as a scaffold[28]. These findings are in accordance with the proposed folding model of the OmpA β-barrel, according to which the protein first binds to the membrane surface in an unfolded, collapsed form and then assumes its tertiary structure through a synchroneous passage of the β-strands and the extracellular loops through the hydrophobic membrane interior[76]. It can be suspected that the inclusion of additives to the refolding buffer should only have a minor effect on the membrane, and are therefore not able to greatly influence the folding process. The nature of the grafted loops and their protonation state, however, can be expected to make a large impact on the folding process as seen from the pH-dependence of refolding. Our results indicate that the high flexibility of the grafted loop sequences at least partially impede the formation of the β-barrel.

3.3 Ligand interaction studies

The primary goals of our grafting attempts were: i) being able to obtain a stably folded chimeric receptor molecule displaying all three extracellular loops of a Y-receptor and ii) obtaining a system in which ligand binding affinity of the grafted loops can be observed, even when outside their natural 7 TM context. The accomplishment of the first aim could be demonstrated for the four three-loop graft constructs based on the analysis of the electrophoretic mobility in SDS-PAGE and the signal dispersion in NMR experiments. In order to test whether our constructs exhibited ligand binding capability, we conducted chemical shift mapping experiments. To this end we assigned all three members of the neurohormones of the NPY family in the presence of DHPC micelles. The titration of ^{15}N-labeled neurohormones with unlabeled receptor construct allowed then the observation of changes in the spectra of the neurohormones, which were interpreted as an interaction taking place between the peptide ligand and the receptor construct. Whereas Y1L1 and Y1L2 had no effect on the spectra of the neurohormones, Y1L3 and Y1L4 induced substantial changes. A quantitative estimation of the interaction strength between the neurohormones and the two receptor constructs is difficult to obtain, as the receptor construct is not conformationally homogenous. *I.e.* it is not present in purely folded form. The folded fraction could only be estimated from comparison of band intensities in SDS-PAGE. Based on the used concentrations in the experiments and the fact that slow exchange occurs in the NMR experiments the dissociation constant can be estimated to be in the medium-to-low micromolar range. The observed changes were the selective disappearance of the C-terminal resonances. The interaction of the neurohormones with DPC micelles has been shown to occur through the C-terminal α-helix, whereas the N-terminal residues are unstructured and pointing away from the micelle surface into the bulk solution[30,32,47]. The C-terminal residues – especially the conserved positively charged arginines at positions 33 and 35 (figure 3) – have been shown to be of particular importance to the

interaction of NPY with the Y1-receptor[58,90,91]. We have therefore produced the two peptides NPY-R33L and NPY-R35L in ^{15}N-labeled form, assigned their spectra in DHPC micelles, and conducted the titration experiments with unlabeled Y1L3. Even though some interaction seems to be retained with these two NPY mutants this interaction is clearly strongly reduced in strength. The specificity of the interaction between the neurohormones and the Y1L3 and Y1L4 receptor constructs is further corroborated by the ability of unlabeled NPY to compete for the binding with ^{15}N-labeled NPY (figure 16) and by the fact that there is no detectable interaction with the Y1L1 and Y1L2 constructs (figure 14).

The complementary experiments, namely the titration of ^{15}N-labeled receptor constructs with unlabeled neurohormones, failed to lead to a detectable effect. This finding is in conflict with the profound changes that the unlabeled Y1L3 and Y1L4 receptor constructs inflicts upon the spectra of the neurohormones. However, careful inspection of the spectra of the receptor constructs reveals that the number of observed sharp resonances, as they are observed from the flexible, grafted loops, is lower than expected. It is therefore possible that the residues responsible for the interaction with the neurohormones are not seen in the spectrum. This can be the case, if they are involved in dynamic processes on the microsecond to millisecond timescale. Such processes are not uncommon for residues at an interface between highly motionally restricted segments as the β-strands of the barrel and highly flexible stretches such as the extracellular loops. While all of the 30 residues of the long e2 loop gave signals in the spectra which could be assigned, in the shorter e1 and e3 loops only 3 out of 15 expected resonances could be observed.

3.4 The influence of loop flexibility on the receptor constructs

Ligand binding of a receptor is always thought as being a balance between favorable enthalpic terms and entropically unfavorable terms, due to the loss of conformational freedom, both of the binding pocket of the receptor and the ligand also. Part of this entropic cost is paid by the pre-arrangement of the interacting residues of the binding pocket. It can be assumed, that if this pre-arrangement is loosened, the interaction strength between the ligand and the receptor decreases. From an NMR point of view, however, the concomitant gain in flexibility could lead to more observable resonances due to the decrease of the exchange broadening phenomenon. We therefore tried to increase the flexibility of the grafted loops - primarily of the e1- and e3-loop sequences - by inserting flexible glycine-serine linkers of different length in between the anchor sites on the scaffold and the Y1-receptor's extracellular loops. Derived from the construct Y1L3 we generated derivatives, carrying glycine-serine linkers of different lenght and flanking different loops: One constructs in which each Y1-receptor loop was flanked at both ends by a Ser-Gly dipeptide and one in which it was flanked by a Ser-Gly-Ser-Gly tetrapeptide (Y1L3-GS and Y1L3-GSGS).

Furthermore two analogous constructs in which only the e1- and e3-loops were flanked, but not the e2-loop, were created (Y1L3-gs and Y1L3-gsgs), and two constructs having only the e2-loop flanked with either the di- or tetrapeptide linker (Y1L3-e2gs and Y1L3-e2gsgs). We observed a significantly different refolding behavior in the different linker-flanked constructs. Only the two constructs carrying the flanking linker sequences at all three loops were able to fold equally well as the parent Y1L3 construct. Whereas the refolding efficiency of these constructs was similar to Y1L3 (~80% efficiency) the two constructs carrying the linker sequences only on the e1- and e3-loop only folded with an efficiency of only ~50%. Constructs carrying the linker sequences only at the e2-loop. These showed the same folding deficiency that was observed for Y1L3-gs and Y1L3-gsgs (figure 20). It seems therefore that the lenghts and maybe also the flexibilties of the loops relative to each other play a role in the folding process of the three-loop graft constructs. However, further experiments would be required to back up this hypothesis.

The [^{15}N,^{1}H]-HSQC spectra of Y1L3-GS and Y1L3-GSGS show additional peaks relative to Y1L3. Most of these, however, stem for the glycine and serine resonances of the linkers themselves, rather than from the intervening loop sequences (figure 20).

Not surprisingly, the interaction of the linker-bearing constructs was slightly weakend compared to Y1L3(figure 20).

3.5 Ion exchange purification of purely folded receptor constructs

We would like to repeat here, that all the interaction studies carried out between the neurohormones and the receptor constructs have been performed with a mixture of folded and unfolded receptor constructs. Even though we had indications, that it is indeed the folded fraction of the construct which is responsible for the interaction with the neurohormones, a definitive proof would only be the repetition of the experiments with 100% folded receptor constructs.

Ion exchange chromatography is a method for separation of molecules based on their charge. Although the charge of the two isoforms of the same protein seems to be identical, differently folded protein species differ in how they display their surface charged groups, which are moreover in different orientation relative to each other, resulting in distinct interaction strengths with the charged groups of an ion-exchange resin. The small column volumes used in ion exchange chromatography makes this technique also appealing to detergent containing buffers.

Even though this technique is applicable to the separation of folded and unfolded wt-OmpA, its efficiency is limited. Both folded and unfolded OmpA elute at very similar salt concentrations, thus giving a bad resolution. The Y1L3 and Y1L4 receptor constructs, in contrast, showed a very clear separation between the unfolded and folded forms. This different behavior can be explained by the increased net number of negatively charged residues (+5) in the grafted loops compared to the

extracellular loops found in the wt-OmpA sequence (figure 7). The signal dispersions encountered in the collected fractions confirms, that they indeed correspond to the folded and unfolded forms of the receptor constructs. The spectrum of the crude mixture can be deconvoluted into spectra from the folded and unfolded forms (figure 21).

The titrations of the neurohormones with this folded form, however, didn't show any sign of interaction (figure 22). Likewise, no interaction between the unfolded fractions and the neurohormones could be observed. Only when all fractions from the purification were mixed, the characteristic decrease in the peak volumes of the C-terminal residues of the neurohormones could be observed. This indicates that for the interaction to occur, both the folded and the unfolded form of the construct have to be present.

4. Materials and methods

4.1 Synthesis and purification of neurohormones

Synthesis of unlabeled neurohormones was carried out using standard Fmoc-based solid-phase peptide synthesis[92] using a robot system (ABI433A, Applied Biosystems). Rink Amide MBHA resin resin was used to assemble the peptides using standard Fmoc chemistry (20% piperidine in DMF for Fmoc deprotection, 4 eq. of HOBt/HBTU for activation, diisopropylthylamine as base, and N-methylpyrrolidone as solvent). Peptides were cleaved from the resin and deprotected with TFA/water/triisopropylsilane/1,2-ethanedithiol/thioanisole in a ratio of 75/4/1/10/10. Cleaved peptide was precipitated and washed twice with ice-cold diethylether. The precipitate was dissolved in water/acetonitrile 8/2, lyophilized and purified by C18 reversed phase HPLC. Peptide masses were confirmed by ESI-MS.

Synthesis of ^{15}N-labeled neurohormones was carried out recombinantly in bacteria as fusions to ubiquitin[93] (for PP) or to ketosteroidisomerase[94] (for NPY, NPY-R33L, NPY-R35L, and PYY). The expression and purification of neurohormones using theses two systems has been described in detail elsewhere[30,47].

4.2 Plasmids of OmpA and its derivatives

The starting plasmid for the transmembrane domain (TMD) of OmpA from *E. coli* was as described in Ramakrishnan *et al.*[95]. Briefly, this plasmid called pTMD-OmpA holds the sequence coding for amino acids 22 to 197 of OmpA as deposited in the UniProtKB/Swiss-Prot database under the

accession number P0A910 (OMPA_ECOLI). A methionine residue replaces the 21 residues periplasmic export signal present in the wild-type sequence. The removal of this signal results in accumulation of TMD-OmpA in the cytoplasm in the form of insoluble inclusion bodies. The coding sequence was cloned into the pET22b plasmid between the *XhoI* and *NdeI* sites. A plasmid map covering the region of the expressed TMD-OmpA sequence can be found in appendix A. All designed constructs were derived from this plasmid and all mutants were created using the QuickChange methodology as described in the QuickChange site directed-mutagenesis kit manual from Stratagene (catalog no. 200518) following a deletion-insertion approach, in which first the original OmpA loop sequence was deleted before the foreign Y1-receptor loop sequence was inserted. All components of the kit were bought individually. The dNTP were from Fermentas (cat. no. #R0241) and were stored as aliquots at -20 °C. *PfuTurbo* polymerase was from Stratagene (cat. no. 600250-52) and *DpnI* nuclease was from Fermentas (cat. no. #ER1701). All QuickChange reactions were carried out in 10 µl total volume containing 1x *PfuTurbo* buffer from Stratagene (cat. no. 600153-82), 0.5 U *PfuTurbo* polymerase, 0.2 mM dNTPs, ~100 ng of plasmid DNA and 0.2 µM of each mutagenic primer. The thermocycling procedure was designed as recommended by Stratagene and can be found in appendix B. 5 U of *DpnI* were added to the reaction and the mixture was incubated at 37 °C for 3 hours. 1-2 µl of this mixture were transformed to 25-50 µl of self-prepared $RbCl_2$-competent cells by heat shock. Plasmid was extracted from the obtained colonies and sequenced by the dideoxy sequencing method [96].

All primers were purchased from Microsynth (Balgach, Switzerland). Primers for deletion, single amino acid change and point mutation reactions were purchased as desalted ~36-mers. Primers for the insertion constructs were self-made by PCR using two short (~12 nucleotides), desalted primers. PCR products were purified with a Sigma PCR clean-up kit (NA1020-1KT) and used in subsequent QuickChange mutagenesis reactions. Since long primers are expensive and require additional expensive purification, this strategy allowed substantial cost savings. The inserted Y-receptor gene sequences were optimized for preferred *E. coli* codon usage.

We followed a strategy in which in a first round all four extracellular loops of the wt-OmpA were removed. The sequences to be deleted were selected based on the construct OmpAΔ1234 studied in by Koebnik[27] and that has served as a scaffold in a previous grafting study[28]. Briefly, in our TMD-OmpA construct[95] residues H19-H31, P62-Y72, K107-G118, and I147-P157 are called OmpA extracellular loops 1, 2, 3, and 4, respectively. In a second round the desired topological arrangement of the Y-receptor loops on the scaffold was generated in four rounds of mutagenesis, filling three positions with Y-receptor loops and the fourth with a minimal turn-inducing sequence as used in the two publications mentioned above. Because of a lack of a high-resolution structure for any of the Y-receptors, for the human Y-receptors (accession number P25929) the extracellular

loops were assumed as annotated in the GPCRDB (http://www.gpcr.org/7tm/), *e.g.* Y100-M112, Q177-S210, and F286-N299 for the Y1-receptor extracellular loops 1, 2, and 3, respectively. Any occurring cysteines in these sequences were replaced by serines. The corresponding DNA sequences were designed using preferred *E. coli* codon usage.

A table with all the primers used to generate the deletions and the insertions can be found in appendix I.

4.3 Expression of OmpA and its derivatives

OmpA and its derivatives were always expressed in unlabeled and in ^{15}N-labeled form. Unlabeled and ^{15}N-labeled proteins were expressed in LB and M9 minimal medium, respectively. The minimal medium contained $^{15}NH_4Cl$ as the sole nitrogen source and its exact composition can be found in appendix C.

A LB preculture containing 100 µg/ml ampicillin was inoculated with a colony of BL21 (DE3) cells carrying the desired plasmid and grown over night at 37 °C and 235 rpm. This preculture was used to inoculate the main culture in a 1:100 dilution. 500 ml of the main culture were grown in 2 l Erlenmeyer flasks at 37 °C and 230 rpm. Cultures were induced with 1 mM IPTG when the OD_{600} reached 0.8. LB cultures were grown for another 2.5-3 h and minimal medium cultures for 4-5 h before harvesting. Cells were harvested by centrifugation at 4 °C for 20 min at 3300 g in a Sorvall GSA rotor. Cell pellets were stored at -20 °C until further processing.

4.4 Expression of ^{2}H,^{13}C,^{15}N-labeled Y1L3

3 ml of D_2O-LB containing 100 µg/ml ampicillin were inoculated with a freshly transformed colony of BL21 (DE3) pY1L3. This preculture was grown at 37 °C and 235 rpm overnight and was used to inoculate 250 ml D_2O-M9 medium containing 0.25 g $^{15}NH_4Cl$ and 1 g ^{2}H,^{13}C-D-glucose. 15 h after inoculation the OD_{600} was 0.9 and the culture was induced with 1 mM IPTG. The culture was grown for another 7 h before it was harvested by centrifugation as described above.

4.5 Purification of OmpA and its derivatives

The purification procedure followed essentially the protocol described in Pautsch *et al.*[68].

The cell pellet was thawed on ice for one hour and resuspended in the 10-fold volume (w/v) cold buffer P (20 mM Tris pH 8.5) by vortexing until no cell clumps were visible any more to yield a milky white solution.

The suspension was sonicated on ice with a Branson digital sonifier for 10 min total sonication time (35% power, 2 s pulse on, 4 s pulse off). The milky white solution was centrifuged at for 90 min at 4300 g on a Sorvall ss-34 rotor at 4 °C.

The bright white pellet was resuspended in 2% Triton X-100 in buffer P by vortexing and with help of a spatula. The solution was centrifuged for 90 min at 4300 g on a Sorvall ss-34 rotor at 4 °C.
The bright white pellet was resuspended in a few milliliters of 8 M urea, 10 mM Tris pH 8, 1 mM EDTA or in 6 M GdnHCl, 10 mM Tris pH 8, 1 mM EDTA as far as possible by vortexing and with help of a spatula (usually for the inclusion bodies from 1 g cell pellet 2 ml of urea solution were used). In case the purified construct contained cysteines 10 mM DTT were added to the urea solution. This solution was incubated at 37 °C and 200 rpm for 2 h. This solution was centrifuged for 20 min at 47800 g at 4 °C on a Sorvall ss-34 rotor. The supernatant was decanted and stored at -20 °C until further use. This procedure typically yielded 5-10 ml solution with a final protein concentration of ~20 mg/ml from 1 liter of bacterial culture.

4.6 Small scale refolding trials with OmpA and its derivatives

Reaction volumes for screening refolding conditions were generally 50 µl unless noted otherwise. The buffers for the refolding screens were 10 mM acetate pH 4, 10 mM HEPES pH 7, 10 mM Tris pH 8.8 and 10 mM borate pH 10 and always contained 1 mM EDTA. In case the refolding constructs contained cysteines an additional 10 mM DTT was added to the refolding solution. The detergent concentration was always chosen to guarantee at least a 500-fold excess of detergent over protein or a concentration double that of the critical micellar concentration (cmc) of the respective detergent. 1.25 µl of a 20 mg/ml protein solution in 8 M urea, 10 mM Tris pH 8, 1 mM EDTA (+ 10 mM DTT if applicable) were added. The solution was mixed by vortexing, briefly spun down and incubated at 30 °C for 5 h. Refolding efficiency was assessed by 18% SDS-PAGE (for exact composition of the gels see appendix D).

4.7 Large scale refolding of OmpA and its derivatives

1 volume of a 20 mg/ml protein solution in 8 M urea, 10 mM Tris pH 8, 1 mM EDTA were diluted at a flow-rate of 1 ml/h with a peristaltic pump into 6 volumes of 3% DHPC, 10 mM borate or piperazine, 1 mM EDTA, pH 10 at 4 °C. The solution was incubated at room temperature for 15 h and dialyzed against a 200-fold excess of cold 50 mM Tris, 5 mM EDTA, pH 8.5 at 4 °C. The solution was then concentrated in an Amicon Ultra-4 centrifugal concentrator (10 kDa MWCO; cat. no. UFC801024) to 250-500 µl and diluted with the 2-fold volume of NMR buffer (3% DHPC, 20 mM sodium phosphate, 100 mM NaCl, 10% D_2O, pH 6.5). This concentration-dilution cycle was repeated once.
The refolding efficiency of the procedure was verified by 18% SDS-PAGE and the concentration of the protein solution was measured by the BCA assay (Thermo Scientific, cat. no. 23227).

4.8 Titration of neurohormones with OmpA and its derivatives

An adequate amount of ^{15}N-labeled neurohormone was dissolved in NMR buffer (3% DHPC, 20 mM sodium phosphate, 100 mM NaCl, 10% D_2O, pH 6.5) and increasing quantities of the refolded receptor constructs were added to reach a final volume of 250 µl. [^{15}N,^{1}H]-HSQC spectra of these samples were measured (for spectroscopic details *vide infra*) and the peak volumes integrated.

4.9 NMR spectroscopy general

All spectra were recorded on a Bruker AV-600 or AV-700 spectrometer both equipped with cryoprobes.

Proton chemical shifts were calibrated to the water signal and nitrogen shifts were referenced indirectly to liquid NH_3[97]. Raw data was processed using the Bruker Topspin software versions 2.0 and 2.1 and transferred to XEASY[98] or CARA[99] for further analysis.

All 2D experiments utilized TPPI-States[100,101] for quadrature detection in indirect proton dimensions, and gradient-selected coherence selection (echo-antiecho)[102] in combination with sensitivity enhancement schemes[102,103] in experiments including detection of amide protons.

4.10 Assignment of NPY, NPY-R33L, NPY-R35L, PYY and PP in DHPC micelles

The reported assignments for NPY[47], PYY[32] and PP[30] in DPC micelles at pH 4.5 served as the starting point for the assignment of the amide resonance of NPY, its two C-terminal Arg-mutants R33L and R35L, PYY and PP in DHPC micelles at pH 6.5. All samples consisted of 0.5 mM ^{15}N-labeled peptide, 3% DHPC, 20 mM sodium phosphate, 100 mM NaCl, 10% D2O, pH 6.5 and spectra were acquired at 310 K with a 1 s interscan delay unless stated otherwise. While some resonances didn't change much and thus could be adopted, others changed too much to allow for a confident assignment without recording further experiments. The adoptable peaks in the [^{15}N,^{1}H]-HSQC served as starting points for the sequential assignment. Neighboring peaks were then linked to these starting points with the help of a [^{1}H,^{1}H]-NOE-relayed [^{15}N,^{1}H]-HSQC[104], which was acquired with a 300 ms mixing time and 2048 and 256 complex points in the direct and indirect dimension, respectively. Water suppression wash achieved through continuous-wave irradiation of the water resonance. A ^{15}N-resolved TOCSY[105] utilizing a 75 ms DIPSI2 mixing sequence was recorded with 2048, 40 and 128 points per increment in the direct and the two indirect (^{15}N, ^{1}H) dimensions respectively. Spin systems were identified in this spectra and used to verify the assignments gained from the NOE-relayed [^{15}N,^{1}H]-HSQC.

4.11 NMR spectroscopy and data analysis of titration of OmpA derivatives with neurohormones

Spectra were recorded on uniformly ^{15}N-labeled receptor construct at concentrations between 0.25 and 1 mM in NMR buffer (3% DHPC, 20 mM sodium phosphate at pH 6.5, 100 mM NaCl and 10% D_2O) and at a temperature of 320 K. Unlabeled ligand was dissolved in a minimal amount of NMR-buffer and added to the receptor construct sample. [^{15}N,^{1}H]-TROSY[106-108] with 1024(^{1}H)*128(^{15}N) complex data points was recorded. Spectral widths were 18 ppm and 36 ppm for the ^{1}H and ^{15}N dimension, respectively. The carrier positions were placed at 4.71 ppm (^{1}H) and 119.5 ppm (^{15}N). 32 scans were recorded per increment.

4.12 Assignment of Y1L3

All spectra for the assignment of Y1L3 were recorded on the AV-700 spectrometer on a 0.5 mM sample of uniformly ^{15}N,^{13}C,^{2}H-labeled Y1L3 in 3% DHPC, 20 mM sodium phosphate at pH 6.5, 100 mM NaCl and 10% D_2O and at a temperature of 320 K.

A [^{15}N,^{1}H]-TROSY[106-108] with 1024(^{1}H)*75(^{15}N) complex data points was recorded. Spectral widths were 18 ppm and 31 ppm for the ^{1}H and ^{15}N dimensions, respectively. The carrier positions were placed at 4.71 ppm (^{1}H) and 119.5 ppm (^{15}N).

A TROSY-version of the HN(CO)CACB[109-111] with 1024(^{1}H)*25(^{15}N)*70(^{13}C) complex data points was recorded. Spectral widths were 18 ppm, 31 ppm and 60 ppm for the ^{1}H, ^{15}N and ^{13}C dimension, respectively. The carrier positions were placed at 4.71 ppm (^{1}H), 118.7 ppm (^{15}N), 39.0 ppm ($^{13}C_{\alpha/\beta}$) and 54 ppm ($^{13}C_{\alpha}$). 16 scans were recorded per increment.

A TROSY-version of the HNCACB[110,111] with 1024(^{1}H)*20(^{15}N)*64(^{13}C) complex data points was recorded. Spectral widths were 18 ppm, 31 ppm and 60 ppm for the ^{1}H, ^{15}N and ^{13}C dimension, respectively. The carrier positions were placed at 4.71 ppm (^{1}H), 118.7 ppm (^{15}N) and 39 ppm (^{13}C). 32 scans were recorded per increment.

A proton-detected version of the steady-state $^{15}N\{^{1}H\}$ heteronuclear Overhauser effect sequence was used for measurement of the heteronuclear NOE[112]. Therein, the buildup of the NOE was achieved through a pulse train of 120 degree proton pulses separated by 5 ms over a period of 3 seconds.

4.13 Calculation of mismatch score

The extracellular loop domain of a GPCR is composed of six anchor points for the loops (*i.e.* the terminal amino acids of the TM helices, which serve as membrane anchors to the extracellular loops). The overall topology of these anchor points is defined by 15 unique distances. Likewise, the topology of the eight anchor points of the four extracellular loops of OmpA is defined by 28 unique distances. Theoretically there are 24 different ways of arranging three foreign loop sequences on the

four acceptor sites of the scaffold. We calculated a "mismatch score" for each of the 24 possible arrangements. The following GPCR crystal structures served to calculate a matrix of the average distances between the anchor points of the extracellular loops: 1) the original crystal structure of bovine rhodopsin[82] (pdb code 1F88), 2) the T4 lysozyme fusion of β2-AR bound to the inverse agonist carazolol[84] (2RH1), 3) the T4 lysozyme fusion of β2-AR bound to the inverse agonist timolol[80] (3D4S), 4) β2-AR in complex the inverse agonist carazolol and bound to a Fab fragment[83] (2R4R), 5) the β1-AR bound to the antagonist cyanopindolol[63] (2VT4), 6) bovine opsin[85] (3CAP), 7) squid rhodopsin[81] (2Z73), 8) squid rhodopsin[113] (2ZIY), 9) the human A_{2A} adenosine receptor in complex with the antagonist ZM241385[114] (3EML), and 10) bovine opsin in complex with a transducin peptide[115] (3DQB).

The distances between the six involved anchor points for each of the possible 24 arrangements of the three Y1-receptor loops on the four acceptor sites of the OmpA scaffold were calculated and compared to the distances calculated for an average GPCR. Mismatches were summed up and interpreted as mismatch scores.

5. References

1. Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr Rev* **21**, 90-113 (2000).
2. Ma, P. & Zemmel, R. Value of novelty? *Nat Rev Drug Discov* **1**, 571-572 (2002).
3. Sollod, B. L. et al. Were arachnids the first to use combinatorial peptide libraries? *Peptides* **26**, 131-139 (2005).
4. Ruiz, N., Kahne, D. & Silhavy, T. J. Advances in understanding bacterial outer-membrane biogenesis. *Nat Rev Microbiol* **4**, 57-66 (2006).
5. Ried, G., Koebnik, R., Hindennach, I., Mutschler, B. & Henning, U. Membrane topology and assembly of the outer membrane protein OmpA of Escherichia coli K12. *Mol Gen Genet* **243**, 127-135 (1994).
6. Pautsch, A. & Schulz, G. E. Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**, 1013-1017 (1998).
7. Pautsch, A. & Schulz, G. E. High-resolution structure of the OmpA membrane domain. *J Mol Biol* **298**, 273-282 (2000).
8. Arora, A., Abildgaard, F., Bushweller, J. H. & Tamm, L. K. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* **8**, 334-338 (2001).
9. Cierpicki, T., Liang, B., Tamm, L. K. & Bushweller, J. H. Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A. *J Am Chem Soc* **128**, 6947-6951 (2006).
10. Koebnik, R., Locher, K. P. & Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37**, 239-253 (2000).
11. Colombini, M. The published 3D structure of the VDAC channel: native or not? *Trends Biochem Sci* **34**, 382-389 (2009).
12. Hiller, S. et al. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206-1210 (2008).
13. Tamm, L. K., Hong, H. & Liang, B. Folding and assembly of beta-barrel membrane proteins. *Biochim Biophys Acta* **1666**, 250-263 (2004).
14. White, S. H., Ladokhin, A. S., Jayasinghe, S. & Hristova, K. How membranes shape protein structure. *J Biol Chem* **276**, 32395-32398 (2001).
15. Killmann, H., Benz, R. & Braun, V. Conversion of the FhuA transport protein into a diffusion channel through the outer membrane of Escherichia coli. *EMBO J* **12**, 3007-3016 (1993).
16. Klebba, P. E., Hofnung, M. & Charbit, A. A model of maltodextrin transport through the sugar-specific porin, LamB, based on deletion analysis. *EMBO J* **13**, 4670-4675 (1994).
17. Charbit, A., Clement, J. M. & Hofnung, M. Further sequence analysis of the phage lambda receptor site. Possible implications for the organization of the lamB protein in Escherichia coli K12. *J Mol Biol* **175**, 395-401 (1984).
18. Killmann, H., Videnov, G., Jung, G., Schwarz, H. & Braun, V. Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. *J Bacteriol* **177**, 694-698 (1995).
19. Morona, R., Klose, M. & Henning, U. Escherichia coli K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. *J Bacteriol* **159**, 570-578 (1984).
20. Beer, K. B. & Miller, V. L. Amino acid substitutions in naturally occurring variants of ail result in altered invasion activity. *J Bacteriol* **174**, 1360-1369 (1992).
21. Kupsch, E. M., Knepper, B., Kuroki, T., Heuer, I. & Meyer, T. F. Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by Neisseria gonorrhoeae for

human leukocytes and epithelial cells. *EMBO J* **12**, 641-650 (1993).
22. Prasadarao, N. V. et al. Outer membrane protein A of Escherichia coli contributes to invasion of brain microvascular endothelial cells. *Infect Immun* **64**, 146-153 (1996).
23. Dornmair, K., Kiefer, H. & Jahnig, F. Refolding of an integral membrane protein. OmpA of Escherichia coli. *J Biol Chem* **265**, 18907-18911 (1990).
24. Surrey, T. & Jähnig, F. Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc Natl Acad Sci U S A* **89**, 7457-7461 (1992).
25. Kleinschmidt, J. H., Wiener, M. C. & Tamm, L. K. Outer membrane protein A of E. coli folds into detergent micelles, but not in the presence of monomeric detergent. *Protein Sci* **8**, 2065-2071 (1999).
26. Schweizer, M., Hindennach, I., Garten, W. & Henning, U. Major proteins of the Escherichia coli outer cell envelope membrane. Interaction of protein II with lipopolysaccharide. *Eur J Biochem* **82**, 211-217 (1978).
27. Koebnik, R. Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from Escherichia coli. *J Bacteriol* **181**, 3688-3694 (1999).
28. Johansson, M. U. et al. A minimal transmembrane beta-barrel platform protein studied by nuclear magnetic resonance. *Biochemistry* **46**, 1128-1140 (2007).
29. Bader, R., Lerch, M. & Zerbe, O. BioNMR in Drug Research. 95-120 (2002).
30. Lerch, M. et al. Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J. Mol. Biol.* **322**, 1117-1133 (2002).
31. Lerch, M. et al. Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding. *Biochemistry* **44**, 9255 - 9264 (2005).
32. Lerch, M., Mayrhofer, M. & Zerbe, O. Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J. Mol. Biol.* **339**, 1153-1168 (2004).
33. Bader, R. & Zerbe, O. Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state? *ChemBioChem* **6**, 1520-1534 (2005).
34. Beck-Sickinger, A. G. & Jung, G. Structure-activity relationships of neuropeptide Y analogues with respect to Y1 and Y2 receptors. *Biopolymers* **37**, 123-142 (1995).
35. Beck-Sickinger, A. G. et al. Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur. J. Biochem.* **225**, 947-958 (1994).
36. Sautel, M. et al. Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. *Mol. Pharmacol.* **50**, 285-292 (1996).
37. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J. Biol. Chem.* **269**, 2863-2869 (1994).
38. Clark, J. T., Kalra, P. S., Crowley, W. R. & Kalra, S. P. Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427-429 (1984).
39. Nakajima, M. et al. Effects of pancreatic polypeptide family peptides on feeding and learning behavior in mice. *J Pharmacol Exp Ther* **268**, 1010-1014 (1994).
40. Larhammar, D. Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11 (1996).
41. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A* **78**, 4175-4179 (1981).
42. Li, X. A., Sutcliffe, M. J., Schwartz, T. W. & Dobson, C. M. Sequence-specific 1H NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **31**, 1245-1253 (1992).

43. Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R. & Zerbe, O. Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin. *Journal of the American Chemical Society* **129**, 8811-8817 (2007).
44. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R. J. Solution structure of monomeric peptide YY supports the functional significance of the PP-fold. *Biochemistry* **39**, 9935-9942 (2000).
45. Cowley, D. J., Hoflack, J. M., Pelton, J. T. & Saudek, V. Structure of neuropeptide Y dimer in solution. *Eur J Biochem* **205**, 1099-1106 (1992).
46. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. Solution structure of human neuropeptide Y. *J Biomol NMR* **8**, 379-390 (1996).
47. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection. *J. Mol. Biol.* **305**, 307-392 (2001).
48. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3**, 842-848 (1996).
49. Herzog, H. et al. Cloned human neuropeptide Y receptor couples to two different second messenger systems. *Proc Natl Acad Sci U S A* **89**, 5794-5798 (1992).
50. Krause, J., Eva, C., Seeburg, P. H. & Sprengel, R. Neuropeptide Y1 subtype pharmacology of a recombinantly expressed neuropeptide receptor. *Mol Pharmacol* **41**, 817-821 (1992).
51. Larhammar, D. et al. Cloning and functional expression of a human neuropeptide Y/peptide YY receptor of the Y1 type. *J Biol Chem* **267**, 10935-10938 (1992).
52. Gehlert, D. R. et al. Expression cloning of a human brain neuropeptide Y Y2 receptor. *Mol Pharmacol* **49**, 224-228 (1996).
53. Gerald, C. et al. Expression cloning and pharmacological characterization of a human hippocampal neuropeptide Y/peptide YY Y2 receptor subtype. *J Biol Chem* **270**, 26758-26761 (1995).
54. Rose, P. M. et al. Cloning and functional expression of a cDNA encoding a human type 2 neuropeptide Y receptor. *J Biol Chem* **270**, 22661-22664 (1995).
55. Bard, J. A., Walker, M. W., Branchek, T. A. & Weinshank, R. L. Cloning and functional expression of a human Y4 subtype receptor for pancreatic polypeptide, neuropeptide Y, and peptide YY. *J Biol Chem* **270**, 26762-26765 (1995).
56. Lundell, I. et al. Cloning of a human receptor of the NPY receptor family with high affinity for pancreatic polypeptide and peptide YY. *J Biol Chem* **270**, 29123-29128 (1995).
57. Gerald, C. et al. A receptor subtype involved in neuropeptide-Y-induced food intake. *Nature* **382**, 168-171 (1996).
58. Cabrele, C. & Beck-Sickinger, A. G. Molecular characterization of the ligand-receptor interaction of the neuropeptide Y family. *J Pept Sci* **6**, 97-122 (2000).
59. Blomqvist, A. G. & Herzog, H. Y-receptor subtypes--how many more? *Trends Neurosci* **20**, 294-298 (1997).
60. Larhammar, D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **65**, 165-174 (1996).
61. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods in Neurosciences* **25**, 366-428 (1995).
62. Walker, P. et al. High level expression of human neuropeptide Y receptors in mammalian cells infected with a recombinant vaccinia virus. *Mol Cell Endocrinol* **91**, 107-112 (1993).
63. Warne, T. et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491 (2008).
64. Gautier, A., Kirkpatrick, J. P. & Nietlispach, D. Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed Engl* **47**, 7297-7300 (2008).

65. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr Opin Biotechnol* **6**, 494-500 (1995).
66. Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol Rev* **60**, 512-538 (1996).
67. Merten, N. et al. Receptor subtype-specific docking of Asp6.59 with C-terminal arginine residues in Y receptor ligands. *J Biol Chem* **282**, 7543-7551 (2007).
68. Pautsch, A., Vogt, J., Model, K., Siebold, C. & Schulz, G. E. Strategy for membrane protein crystallization exemplified with OmpA and OmpX. *Proteins* **34**, 167-172 (1999).
69. Le Maire, M., Champeil, P. & Møller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *BBA-Biomembranes* (2000).
70. Garavito, R. M., Hinz, U. & Neuhaus, J. M. The crystallization of outer membrane proteins from Escherichia coli. Studies on lamB and ompA gene products. *J Biol Chem* **259**, 4254-4257 (1984).
71. Lauterwein, J., Bosch, C., Brown, L. R. & Wüthrich, K. Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. *Biochim Biophys Acta* **556**, 244-264 (1979).
72. Chou, J. J., Baber, J. L. & Bax, A. Characterization of phospholipid mixed micelles by translational diffusion. *J Biomol NMR* **29**, 299-308 (2004).
73. Lin, T. L., Chen, S. H., Gabriel, N. E. & Roberts, M. F. The use of small-angle neutron scattering to determine the structure and interaction of dihexanoylphosphatidylcholine micelles. *J Am Chem Soc* **108**, 3499-3507 (1986).
74. Sanders, C. R. & Sonnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-40 (2006).
75. Fernandez, C., Adeishvili, K. & Wüthrich, K. Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. *Proc Natl Acad Sci U S A* **98**, 2358-2363 (2001).
76. Kleinschmidt, J. H., den Blaauwen, T., Driessen, A. J. & Tamm, L. K. Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **38**, 5006-5016 (1999).
77. Kleinschmidt, J. H. & Tamm, L. K. Time-resolved distance determination by tryptophan fluorescence quenching: probing intermediates in membrane protein folding. *Biochemistry* **38**, 4996-5005 (1999).
78. Larhammar, D. & Salaneck, E. Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-151 (2004).
79. Mobarec, J. C., Sanchez, R. & Filizola, M. Modern Homology Modeling of G-Protein Coupled Receptors: Which Structural Template to Use? *J Med Chem* **52**, 5207-5216 (2009).
80. Hanson, M. A. et al. A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure* **16**, 897-905 (2008).
81. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363-367 (2008).
82. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
83. Rasmussen, S. G. et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383-387 (2007).
84. Cherezov, V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
85. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183-187 (2008).
86. Han, S. J. et al. Identification of an agonist-induced conformational change occurring adjacent to the ligand-binding pocket of the M(3) muscarinic acetylcholine receptor. *J Biol Chem* **280**, 34849-34858 (2005).
87. Mijares, A., Lebesgue, D., Wallukat, G. & Hoebeke, J. From agonist to antagonist: Fab

fragments of an agonist-like monoclonal anti-beta(2)-adrenoceptor antibody behave as antagonists. *Mol Pharmacol* **58**, 373-379 (2000).
88. Rose, P. M. et al. Aspartate mutation distinguishes ETA but not ETB receptor subtype-selective ligand binding while abolishing phospholipase C activation in both receptors. *FEBS Lett* **361**, 243-249 (1995).
89. Shi, L. & Javitch, J. A. The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. *Proc Natl Acad Sci U S A* **101**, 440-445 (2004).
90. Kirby, D. A., Boublik, J. H. & Rivier, J. E. Neuropeptide Y: Y1 and Y2 affinities of the complete series of analogues with single D-residue substitutions. *J Med Chem* **36**, 3802-3808 (1993).
91. Lindner, D., Stichel, J. & Beck-Sickinger, A. G. Molecular recognition of the NPY hormone family by their receptors. *Nutrition* **24**, 907-917 (2008).
92. Merrifield, B. Concept and early development of solid-phase peptide synthesis. *Methods Enzymol* **289**, 3-13 (1997).
93. Kohno, T., Kusunoki, H., Sato, K. & Wakamatsu, K. A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and ^{15}N/^{13}C. *J Biomol NMR* **12**, 109-121 (1998).
94. Kuliopulos, A. & Walsh, C. Production, Purification, and Cleavage of Tandem Repeats of Recombinant Peptides. *J. Am. Chem. Soc,* **116**, 4599-4607 (1994).
95. Ramakrishnan, M., Qu, J., Pocanschi, C. L., Kleinschmidt, J. H. & Marsh, D. Orientation of beta-barrel proteins OmpA and FhuA in lipid membranes. Chain length dependence from infrared dichroism. *Biochemistry* **44**, 3515-3523 (2005).
96. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
97. Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. Observation of 1000-fold enhancement of nitrogen-15 NMR via proton-detected multiquantum coherences: studies of large peptides. *J Am Chem Soc* **106**, 6104-6105 (1984).
98. Bartels, C., Xia, T. H., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **5**, 1-10 (1995).
99. Keller, R. The Computer Aided Resonance Assignment. (2004).
100. Bodenhausen, G., Vold, R. L. & Vold, R. R. Multiple quantum spin-echo spectroscopy. *J Magn Reson* **37**, 93-106 (1980).
101. Marion, D. & Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun* **113**, 967-974 (1983).
102. Kay, L. E., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc,* **114**, 10663-10665 (1992).
103. Palmer, A. G. I. I. I., Cavanagh, J., Wright, P. E. & Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson* **93**, 151-170 (1991).
104. Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. A powerful method of sequential proton resonance assignment in proteins using relayed 15N-1H multiple quantum coherence spectroscopy. *FEBS Lett* **243**, 93-98 (1989).
105. Bax, A. D. & Davis, D. G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. magn. Reson* **65**, 355 (1985).
106. Pervushin, K. V., Wider, G. & Wüthrich, K. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *Journal of Biomolecular NMR* **12**, 345-348 (1998).
107. Rance, M., Loria, J. P. & Palmer, A. G. Sensitivity improvement of transverse relaxation-

optimized spectroscopy. *Journal of Magnetic Resonance* (1999).
108. Zhu, G., Kong, X. M. & Sze, K. H. Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D. *Journal of Biomolecular NMR* (1999).
109. Dötsch, V., Matsuo, H. & Wagner, G. Amino-acid-type identification for deuterated proteins with a β-carbon-edited. *Journal of Magnetic Resonance* (1996).
110. Eletsky, A., Kienhofer, A. & Pervushin, K. TROSY NMR with partially deuterated proteins. *J Biomol NMR* **20**, 177-180 (2001).
111. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc* **121**, 844 (1999).
112. Noggle, J. H. & Schirmer, R. E. The Nuclear Overhauser Effect - Chemical Applications. (1971).
113. Shimamura, T. et al. Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J Biol Chem* **283**, 17753-17756 (2008).
114. Jaakola, V. P. et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211-1217 (2008).
115. Scheerer, P. et al. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497-502 (2008).

6. Appendix

A - plasmid map covering the region of the expressed TMD-OmpA sequence (bold letters)

```
NdeI restriction site:  5' c a't a t g 3'
                        3' g t a t'a c 5'

XhoI restriction site:  5' c't c g a g 3'
                        3' g a g c t'c 5'

cgccatatggctccgaaagataacacctggtacactggtgccaaactgggctggtctcag
 R  H  M  A  P  K  D  N  T  W  Y  T  G  A  K  L  G  W  S  Q
taccatgacactggtttcatcaacaacaatggcccgacccatgaaaaccaactgggcgct
 Y  H  D  T  G  F  I  N  N  N  G  P  T  H  E  N  Q  L  G  A
ggtgcttttggtggttaccaggttaacccgtatgttggctttgaaatgggttacgaatgg
 G  A  F  G  G  Y  Q  V  N  P  Y  V  G  F  E  M  G  Y  E  W
ttaggtcgtatgccgtacaaaggcagcgttgaaaacggtgcatacaaagctcagggcgtt
 L  G  R  M  P  Y  K  G  S  V  E  N  G  A  Y  K  A  Q  G  V
caactgaccgctaaactgggttacccaatcactgacgacctggacatctacactcgtctg
 Q  L  T  A  K  L  G  Y  P  I  T  D  D  L  D  I  Y  T  R  L
ggtggcatggtatggcgggccgacactaaatccaacgtatacggtaaaaaccacgacacc
 G  G  M  V  W  R  A  D  T  K  S  N  V  Y  G  K  N  H  D  T
ggcgtttctccggtcttcgctggcggtgttgagtacgcgatcactcctgaaatcgctacc
 G  V  S  P  V  F  A  G  G  V  E  Y  A  I  T  P  E  I  A  T
cgtctagaataccagtggaccaacaacatcggtgacgcacacaccatcggcactcgtccg
 R  L  E  Y  Q  W  T  N  N  I  G  D  A  H  T  I  G  T  R  P
gacaacggcatgctgagcctgggtgtttcctaccgtttcggtcagggcgaagcagcttga
 D  N  G  M  L  S  L  G  V  S  Y  R  F  G  Q  G  E  A  A  -
ctcgagcag
 L  E  Q
```

B - Thermocycling for QuickChange mutagenesis

95 °C	3 min
---	---
95 °C	30 s
55 °C	1 min
68 °C	6 min
---	---
4 °C	∞

The central block was repeated 18 times for insertions and deletions, 16 times for single amino acid changes and 12 times for point mutations.

C - Composition of M9 minimal medium

Amounts are for 1 l medium:

4 g KH_2PO_4

4 g K_2HPO_4

3.5 g $Na_2HPO_4 * 2\ H_2O$

1 g NaCl

1 g $^{15}NH_4Cl$ (or $^{14}NH_4Cl$ if no labeling was required)

fill up to 960 ml with de-ionized water and autoclave

To these 960 ml the following were added:

10 ml $MgSO_4$ 1 M (autoclaved)

25 ml glucose 20% w/v (sterile filtered)

2 ml trace metal stock solution (autoclaved)

1 ml thiamineHCl 150 mM (sterile filtered)

1 ml ampicillin 100 mg/ml

trace metal stock solution (500X):

$FeSO_4 * 7\ H_2O$	4 g/l
$CaCl_2 * 2\ H_2O$	4 g/l
$AlCl_3 * 6\ H_2O$	1 g/l
$MnSO_4 * n\ H_2O$	1 g/l
$CoCl_2 * 6\ H_2O$	0.4 g/l
$ZnSO_4 * 7\ H_2O$	0.2 g/l
$CuCl_2 * 2\ H_2O$	0.1 g/l
H_3BO_3	0.1 g/l

D - 18% SDS-PAGE for assaying the folding state of OmpA and its derivatives

The stacking gel was prepared as described in Sambrook et al. (Sambrook J et al. 2000. Molecular Cloning: A Laboratory Manual (Third Edition)).

The resolving gel was composed as follows (for 10 ml of resolving gel):

2.5 ml 1.5 M Tris pH 8

6 ml 30% acrylamide

1.3 ml ddH2O

0.1 ml 10% SDS

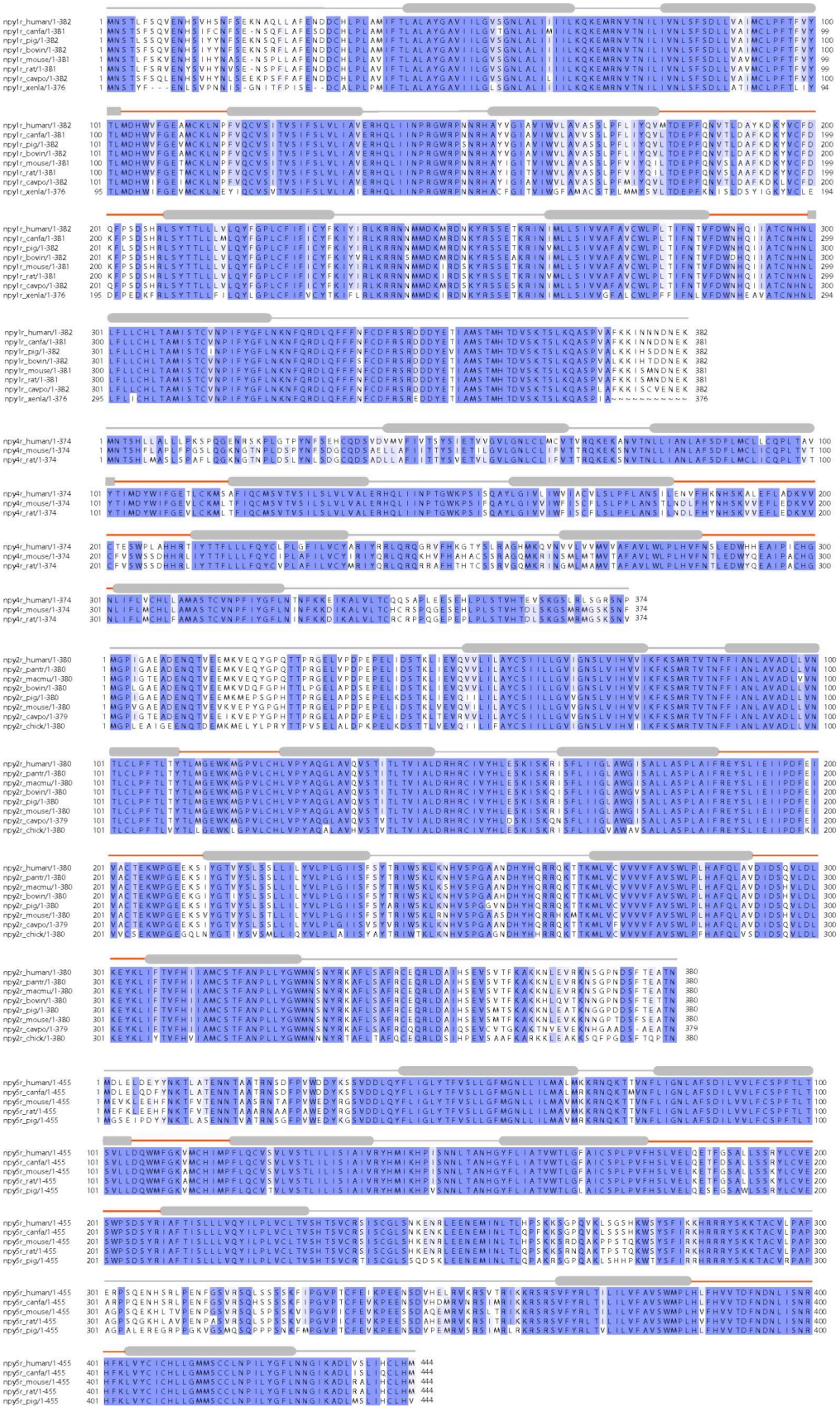
0.1 ml 10% APS

0.017 ml TEMED

E - GPCR structures deposited in the Protein Data Bank as of December 2009

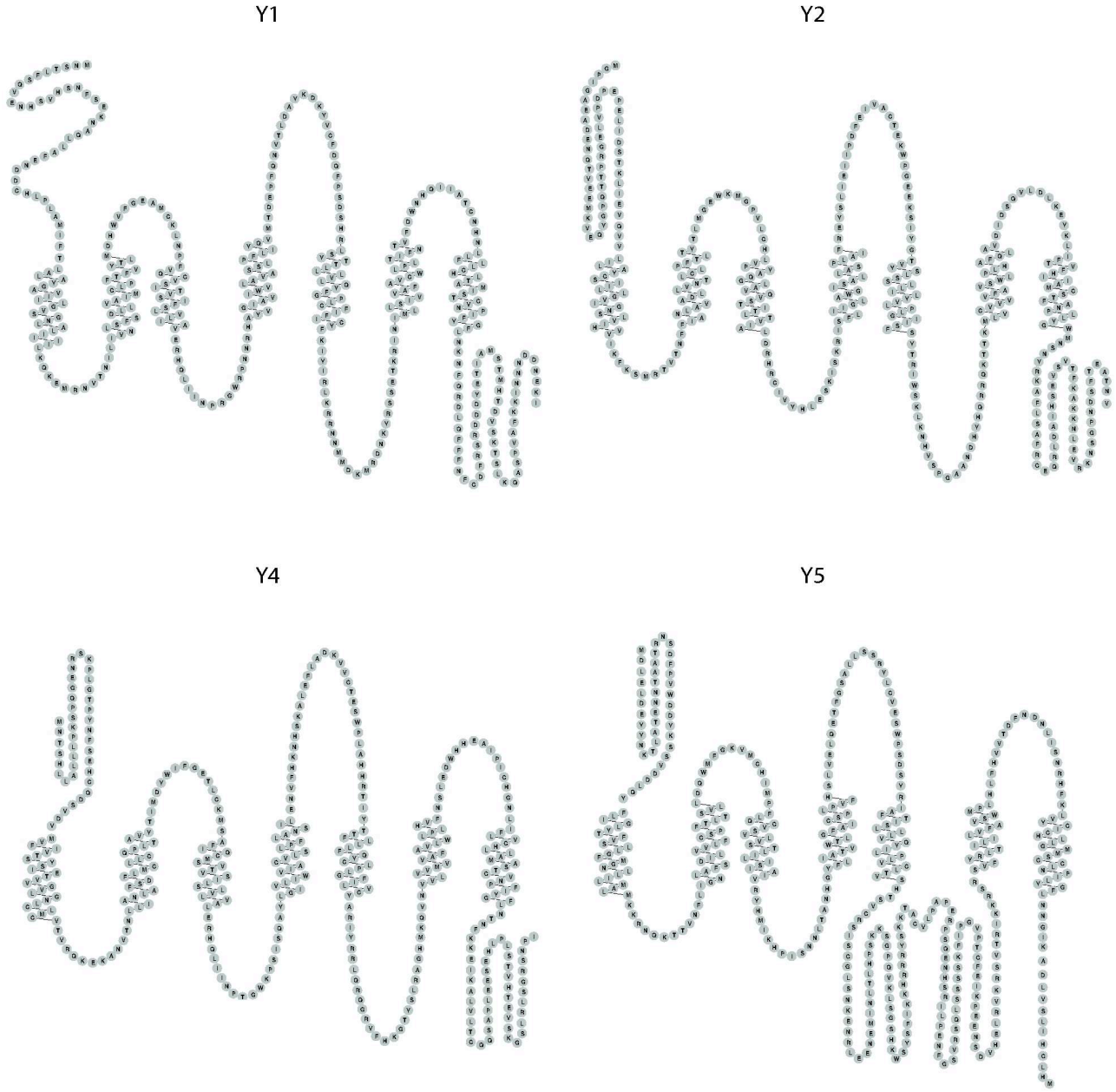
GPCR	Description	pdb
Bovine rhodopsin	First crystal structure of rhodopsin	1F88
Bovine rhodopsin	Alternative model of ground state rhodopsin deposited under pdb code 1GZM	3C9L
Bovine rhodopsin N2C/D282C mutant	Alternative model of the N2C/D282C thermostable rhodopsin mutant deposited under pdb code 2J4Y	3C9M
Bovine rhodopsin	Rhodopsin in trigonal crystal form	1GZM
Bovine rhodopsin	Study of the role of internal water molecules in the rhodopsin crystal structure	1L9H
Bovine rhodopsin	Refinement of rhodopsin crystal structure with focus on the retinal conformation	1U19
Bovine rhodopsin	Refinement of rhodopsin crystal structure	1HZX
Bovine rhodopsin N2C/D282C mutant	Crystal structure of a thermally stable N2C/D282C rhodopsin mutant	2J4Y
Bovine rhodopsin 9-cis form	Crystal structure of 9-cis rhodopsin and comparison with 11-cis rhodopsin and bathorhodopsin	2PED
Bovine rhodopsin lumirhodopsin	Crystal structure of lumirhodopsin (=nearly all-trans rhodopsin)	2HPY
Bovine rhodopsin bathorhodospin	Crystal structure of bathorhodopsin, the first activation intermediate	2G87
Bovine rhodopsin	Crystal structure of rhombohedral crystal form of ground-state rhodopsin	2I35
Bovine rhodopsin	Crystal structure of trigonal crystal form of ground-state rhodopsin	2I36
Bovine rhodopsin	Crystal structure of photoactivated deprotonated intermediate of rhodopsin, reminiscent of metarhodopsin II the G-protein activating state	2I37
Human β2-AR Fab complex	Crystal structure of human β2-AR in complex with a Fab fragment; in complex with the inverse agonist carazolol	2R4R
Human β2-AR Fab complex	Crystal structure of human β2-AR with an engineered TEV cleavage site after residue 24 in the N-terminus in complex with a Fab fragment; in complex with the inverse agonist carazolol	2R4S
Human β2-AR T4L fusion	Crystal structure of human β2-AR carrying the T4 lysozyme sequence in its i3 loop; in complex with the inverse agonist carazolol	2RH1
Human β2-AR T4L fusion	Crystal structure of β2-AR-T4L in complex with the inverse agonist timolol revealing the binding sites for two cholesterol molecules	3D4S
Squid rhodopsin	Crystal structure of squid rhodopsin	2Z73
Human β1-AR	Crystal structure of human β1-AR thermostabilized through point mutations in complex with the antagonist cyanopindolol	2VT4
Human A2A-adenosine receptor	Crystal structure of the human A2A-adenosine receptor in complex with the antagonist ZM241385	3EML
Bovine opsin	Crystal structure of bovine opsin	3CAP
BovineOpsin+GaCT	Crystal structure of bovine opsin in complex with transducin peptide	3DQB
Squid rhodopsin	Crystal structure of squid rhodopsin with extended cytoplasmic region	2ZIY

F - Multiple sequence alignments of the four Y-receptors from a number of major mammalian species. The intensity of the blue coloring indicates the conservation levels of different residues.



Above the sequence the predicted TM helices are indicated by gray rods and the intracellular loops and N- and C-termini by gray lines. Extracellular loops are represented by red lines.

G - Snake plot representations of the four human Y-receptors.



H - Amide proton and nitrogen resonance assignments of pNPY, pPYY, bPP, pNPY-R33L and pNPY-R35L in 3% DHPC, 20 mM sodium phosphate, 100 mM NaCl at pH 6.5 and 37 °C.

pNPY

res.	no.	shift	atom
S	3	8.402	H
		116.774	N
K	4	8.277	H
		123.623	N
D	6	8.238	H
		118.872	N
N	7	8.183	H
		118.644	N
G	9	8.395	H
		108.663	N
E	10	8.102	H
		120.229	N
D	11	8.375	H
		121.116	N
A	12	8.151	H
		125.048	N
A	14	8.399	H
		123.798	N
E	15	8.464	H
		119.449	N
D	16	8.188	H
		120.729	N
L	17	8.203	H
		121.237	N
A	18	8.098	H
		121.885	N
R	19	7.984	H
		118.812	N
Y	20	7.988	H
		119.128	N
Y	21	8.417	H
		118.862	N
S	22	8.361	H
		114.289	N
A	23	8.003	H
		124.509	N
L	24	8.291	H
		118.649	N
R	25	8.329	H
		118.260	N
H	26	7.978	H
		118.339	N

pNPY

res.	no.	shift	atom
Y	27	8.308	H
		119.048	N
I	28	8.580	H
		118.310	N
N	29	8.253	H
		120.294	N
L	30	7.876	H
		120.021	N
I	31	8.043	H
		116.427	N
T	32	8.044	H
		112.523	N
R	33	7.862	H
		120.687	N
Q	34	7.996	H
		117.723	N
R	35	8.092	H
		120.717	N
Y	36	7.985	H
		119.922	N

pNPY-R33L

res.	no.	shift	atom
S	3	8.388	H
		116.675	N
K	4	8.278	H
		123.636	N
D	6	8.247	H
		118.901	N
N	7	8.201	H
		118.666	N
G	9	8.393	H
		108.732	N
E	10	8.102	H
		120.250	N
D	11	8.376	H
		121.154	N
A	12	8.155	H
		125.071	N
A	14	8.408	H
		123.966	N
E	15	8.489	H
		119.575	N
D	16	8.178	H
		120.809	N
L	17	8.198	H
		121.309	N
A	18	8.085	H
		120.718	N
R	19	7.959	H
		118.259	N
Y	20	7.957	H
		119.819	N
Y	21	8.377	H
		118.772	N
S	22	8.335	H
		114.225	N
A	23	7.959	H
		124.024	N
L	24	8.226	H
		118.226	N
R	25	8.248	H
		118.183	N
H	26	8.014	H
		117.740	N

pNPY-R33L

res.	no.	shift	atom
Y	27	8.280	H
		118.757	N
I	28	8.486	H
		118.083	N
N	29	8.243	H
		120.479	N
L	30	8.020	H
		119.991	N
I	31	8.063	H
		116.799	N
T	32	7.999	H
		114.744	N
L	33	7.956	H
		121.527	N
Q	34	8.020	H
		117.096	N
R	35	7.903	H
		117.839	N
Y	36	7.923	H
		118.671	N

pNPY-R35L

res.	no.	shift	atom
S	3	8.403	H
		116.766	N
K	4	8.286	H
		123.661	N
D	6	8.252	H
		118.950	N
N	7	8.196	H
		118.649	N
G	9	8.390	H
		108.685	N
E	10	8.101	H
		120.148	N
D	11	8.374	H
		121.037	N
A	12	8.136	H
		125.029	N
A	14	8.409	H
		123.813	N
E	15	8.480	H
		119.360	N
D	16	8.162	H
		120.761	N
L	17	8.190	H
		121.416	N
A	18	8.089	H
		120.744	N
R	19	7.972	H
		118.376	N
Y	20	7.995	H
		120.218	N
Y	21	8.448	H
		118.826	N
S	22	8.357	H
		114.281	N
A	23	8.000	H
		124.781	N
L	24	8.340	H
		118.778	N
R	25	8.318	H
		118.699	N
H	26	8.344	H
		118.076	N

pNPY-R35L

res.	no.	shift	atom
Y	27	8.051	H
		117.667	N
I	28	8.626	H
		118.689	N
N	29	8.296	H
		120.622	N
L	30	7.866	H
		120.148	N
I	31	8.150	H
		117.721	N
T	32	8.334	H
		113.353	N
R	33	7.890	H
		120.295	N
Q	34	7.792	H
		117.121	N
L	35	7.949	H
		118.158	N
Y	36	7.805	H
		116.520	N

pPYY

res.	no.	shift	atom
A	3	8.364	H
		124.446	N
K	4	8.264	H
		122.200	N
E	6	8.508	H
		121.937	N
A	7	8.330	H
		126.541	N
G	9	8.326	H
		108.002	N
E	10	8.587	H
		118.801	N
D	11	8.299	H
		118.757	N
A	12	8.488	H
		124.135	N
S	13	8.246	H
		116.150	N
E	15	8.391	H
		119.362	N
E	16	8.528	H
		121.462	N
L	17	7.966	H
		118.932	N
S	18	8.270	H
		114.011	N
R	19	8.596	H
		117.970	N
Y	20	8.257	H
		119.318	N
Y	21	8.462	H
		119.219	N
A	22	7.671	H
		123.147	N
S	23	8.586	H
		118.145	N
L	24	8.305	H
		120.673	N
R	25	7.929	H
		121.958	N
H	26	8.057	H
		121.404	N

pPYY

res.	no.	shift	atom
Y	27	8.560	H
		119.063	N
L	28	8.167	H
		121.131	N
N	29	8.155	H
		118.929	N
L	30	7.823	H
		119.776	N
V	31	8.089	H
		115.932	N
T	32	7.826	H
		112.592	N
R	33	7.855	H
		120.566	N
Q	34	8.004	H
		117.879	N
R	35	8.027	H
		121.235	N
Y	36	8.008	H
		118.961	N

bPP

res.	no.	shift	atom
L	3	8.309	H
		122.211	N
E	4	8.163	H
		122.805	N
E	6	8.454	H
		121.947	N
Y	7	8.448	H
		125.215	N
G	9	8.405	H
		111.365	N
D	10	8.272	H
		117.062	N
N	11	8.242	H
		118.943	N
A	12	7.563	H
		122.855	N
T	13	8.267	H
		113.118	N
E	15	8.645	H
		118.242	N
Q	16	7.871	H
		120.371	N
M	17	8.379	H
		118.890	N
A	18	8.190	H
		122.007	N
Q	19	7.845	H
		119.403	N
Y	20	8.067	H
		120.298	N
A	21	8.378	H
		120.761	N
A	22	7.813	H
		119.650	N
E	23	8.165	H
		119.897	N
L	24	8.591	H
		121.440	N
R	25	8.108	H
		118.230	N
R	26	7.698	H
		118.230	N

bPP

res.	no.	shift	atom
Y	27	8.144	H
		120.298	N
I	28	8.480	H
		118.261	N
N	29	8.153	H
		118.971	N
M	30	7.825	H
		118.446	N
L	31	7.694	H
		119.042	N
T	32	7.591	H
		107.722	N
R	33	7.630	H
		123.168	N
R	35	8.195	H
		120.231	N
Y	36	7.923	H
		120.205	N

I - Primers used in the Quick Change mutagenesis reactions

Primer name	5'->3' sequence	Synthesis	Purification	Comments
Y1e1.fw	tacaccctgatggaccactgggttttcggtgaagcgatg	solid phase	desalted	sense-strand coding for the first extracellular loop of the Y1-receptor (Y1e1) optimized for *E. coli* codon usage
Y1e1.rv	catcgcttcaccgaaaacccagtggtccatcagggtgta	solid phase	desalted	antisense-strand coding for the first extracellular loop of the Y1-receptor (Y1e1) optimized for *E. coli* codon usage
Y1e2C197S.fw	caggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttctttt	solid phase	PAGE	sense-strand coding for the second extracellular loop of the Y1-receptor carrying a C197S mutation (Y1e2C197S) optimized for *E. coli* codon usage
Y1e2C197S.rv	agacagacggtgagaatcggacgggaactgatcaaaagaaacgtatttgtctttgtacgcatccagggt	solid phase	PAGE	antisense-strand coding for the second extracellular loop of the Y1-receptor carrying a C197S mutation (Y1e2C197S) optimized for *E. coli* codon usage
Y1e2C197S.ds	caggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttcttttgatcagttcccgtccgattctcaccgtctgtct	PCR	spin column	double-stranded sequence coding for the second extracellular loop of the Y1-receptor carrying a C197S mutation (Y1e2C197S) optimized for *E. coli* codon usage
Y1e2.fw	aaagacaaatacgtttgctttgatcagttcccgtcc	solid phase	desalted	forward primer for mutating the S197 in the Y1e2C197 back to the original C
Y1e2.rv	ggacgggaactgatcaaagcaaacgtatttgtcttt	solid phase	desalted	reverse primer for mutating the S197 in the Y1e2C197 back to the original C
Y1e3C294S.fw	ttcgactggaaccaccagatcatcgcgacctctaaccacaac	solid phase	desalted	sense-srand coding for the third extracellular loop of the Y1-receptor carrying a C294S mutation (Y1e3C294S) optimized for *E. coli* codon usage
Y1e3C294S.rv	gttgtggttagaggtcgcgatgatctggtggttccagtcgaa	solid phase	desalted	antisense-strand coding for the third extracellular loop of the Y1-receptor carrying a C294S mutation (Y1e3C294S) optimized for *E. coli* codon usage
d1.fw	ctgggctggtctcagtacgaaaaccaactgggcgctgg	solid phase	desalted	sense-strand of the deletion primer for the removal of the first extracellular loop of OmpA
d1.rv	ccagcgcccagttggttttcgtactgagaccagcccag	solid phase	desalted	antisense-strand of the deletion primer for the removal of the first extracellular loop of OmpA
d2.fw	gaatggttaggtcgtatgaaagctcagggcgttcaac	solid phase	desalted	sense-strand of the deletion primer for the removal of the second extracellular loop of OmpA
d2.rv	gttgaacgccctgagctttcatacgacctaaccattc	solid phase	desalted	antisense-strand of the deletion primer for the removal of the fsecond extracellular loop of OmpA
d3.fw	gtatggcgggccgacactgtttctccggtcttcgc	solid phase	desalted	sense-strand of the deletion primer for the removal of the third extracellular loop of OmpA
d3.rv	gcgaagaccggagaaacagtgtcggcccgccatac	solid phase	desalted	antisense-strand of the deletion primer for the removal of the third extracellular loop of OmpA
d4.fw	gaataccagtggaccaacaacgacaacggcatgctgagc	solid phase	desalted	sense-strand of the deletion primer for the removal of the fourth extracellular loop of OmpA
d4.rv	gctcagcatgccgttgtcgttgttggtccactggtattc	solid phase	desalted	antisense-strand of the deletion primer for the removal of the fourth extracellular loop of OmpA
L1.fw	ggctggtctcagtactctcgtgaaaaccaactgggc	solid phase	desalted	sense-strand of the insertion primer for the introduction of the SR-linker replacing the first extracellular loop of OmpA
L1.rv	gcccagttggttttcacgagagtactgagaccagcc	solid phase	desalted	antisense-strand of the insertion primer for the introduction of the SR-linker replacing the first extracellular loop of OmpA
L3.fw	tggcgggccgacactctctgtttctccggtcttcg	solid phase	desalted	sense-strand of the insertion primer for the introduction of the S-linker replacing the third extracellular loop of OmpA
L3.rv	cgaagaccggagaaacagaagtgtcggcccgcca	solid phase	desalted	antisense-strand of the insertion primer for the introduction of the S-linker replacing the third extracellular loop of OmpA
L4.fw	cagtggaccaacaacgcgtctgacaacggcatgctg	solid phase	desalted	sense-strand of the insertion primer for the introduction of the AS-linker replacing the fourth extracellular loop of OmpA
L4.rv	cagcatgccgttgtcagacgcgttgttggtccactg	solid phase	desalted	antisense-strand of the insertion primer for the introduction of the AS-linker replacing the fourth extracellular loop of OmpA

d1Y1e1.fw	ctgggctggtctcagtactacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e1.rv	gcgcccagttggttttccatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e1.ds	ctgggctggtctcagtactacaccctgatggaccactgggttttcggtgaagcgatggaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1 sequence replacing the first extracellular loop of OmpA
d1Y1e1CK.fw	d1Y1e1.fw	solid phase	desalted	forward primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e1CK.rv	gcgcccagttggttttctttgcacatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e1CK.ds	ctgggctggtctcagtactacaccctgatggaccactgggttttcggtgaagcgatgtgcaaagaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1CK sequence replacing the first extracellular loop of OmpA
d2Y1e1.fw	cgaatggttaggtcgtatgtacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1.rv	cagttgaacgccctgagctttcatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1.ds	cgaatggttaggtcgtatgtacaccctgatggaccactgggttttcggtgaagcgatgaaagctcagggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1 sequence replacing the second extracellular loop of OmpA
d2Y1e1CK.fw	d2Y1e1.fw	solid phase	desalted	forward primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1CK.rv	cagttgaacgccctgagctttttgcacatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1CK.ds	cgaatggttaggtcgtatgtacaccctgatggaccactgggttttcggtgaagcgatgtgcaaaaaagctcagggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1CK sequence replacing the second extracellular loop of OmpA
d2SGY1e1.fw	cgaatggttaggtcgtatgtctggttacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with a SG-linker and the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1GS.rv	cagttgaacgccctgagctttggaaccatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with a GS-linker and the sequence adjacent to the second extracellular loop of OmpA
d2SGY1e1GS.ds	cgaatggttaggtcgtatgtctggttacaccctgatggaccactgggttttcggtgaagcgatggttccaaaagctcaggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the SGY1e1GS sequence replacing the second extracellular loop of OmpA
d2SGSGY1e1.fw	cgaatggttaggtcgtatgtctggttccggctacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with a SGSG-linker and the sequence adjacent to the second extracellular loop of OmpA
d2Y1e1GSGS.rv	cagttgaacgccctgagctttgctgccggaaccatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with a GSGS-linker and the sequence adjacent to the second extracellular loop of OmpA
d2SGSGY1e1GSGS.ds	cgaatggttaggtcgtatgtctggttccggctacaccctgatggaccactgggttttcggtgaagcgatggttccggcagcaaagctcagggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the SGSGY1e1GSGS sequence replacing the second extracellular loop of OmpA
d3Y1e1.fw	gtatggcgggccgacactacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e1.rv	gcgaagaccggagaaaccatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e1.ds	gtatggcgggccgacactacaccctgatggaccactgggttttcggtgaagcgatggtttctccggtcttcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1 sequence replacing the third extracellular loop of OmpA
d3Y1e1CK.fw	d3Y1e1.fw	solid phase	desalted	forward primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e1CK.rv	gcgaagaccggagaaacttttgcacatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e1CK.ds	gtatggcgggccgacactacaccctgatggaccactgggttttcggtgaagcgatgtgcaaagtttctccggtcttcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1CK sequence replacing the third extracellular loop of OmpA

d4Y1e1.fw	ccagtggaccaacaactacaccctgatggacc	solid phase	desalted	forward primer for flanking the Y1e1 sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e1.rv	gctcagcatgccgttgtccatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e1.ds	ccagtggaccaacaactacaccctgatggaccactgggttttcggtgaaagcgatggacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1 sequence replacing the fourth extracellular loop of OmpA
d4Y1e1CK.fw	d4Y1e1.fw	solid phase	desalted	forward primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e1CK.rv	gctcagcatgccgttgtctttgcacatcgcttcaccgaaa	solid phase	desalted	reverse primer for flanking the Y1e1 sequence with an additional CK with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e1CK.ds	ccagtggaccaacaactacaccctgatggaccactgggttttcggtgaaagcgatgtgcaaagacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e1CK sequence replacing the fourth extracellular loop of OmpA
d1Y1e2C197S.fw	ctgggctggtctcagtaccaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e2C197S.rv	gcgcccagttggttttcagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e2C197S.ds	ctgggctggtctcagtaccaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgttttcttttgatcagttcccgtccgattctcaccgtctgtctgaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e2C197S sequence replacing the first extracellular loop of OmpA
d2Y1e2C197S.fw	cgaatggttaggtcgtatgcaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e2C197S.rv	cagttgaacgccctgagctttagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e2C197S.ds	cgaatggttaggtcgtatgcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttcttttgatcagttcccgtccgattctcaccgtctgtctaaagctcagggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e2C197S sequence replacing the second extracellular loop of OmpA
d3Y1e2C197S.fw	gtatggcgggccgacactcaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e2C197S.rv	gcgaagaccggagaaacagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e2C197S.ds	gtatggcgggccgacactcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgttttcttttgatcagttcccgtccgattctcaccgtctgtctgtttctccggtcttcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e2C197S sequence replacing the third extracellular loop of OmpA
d4Y1e2C197S.fw	ccagtggaccaacaaccaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e2C197S.rv	gctcagcatgccgttgtcagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e2C197S.ds	ccagtggaccaacaaccaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttcttttgatcagttcccgtccgattctcaccgtctgtctgacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e2C197S sequence replacing the fourth extracellular loop of OmpA
d4SGY1e2C197S.fw	ccagtggaccaacaactctggtcaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with a SG-linker and the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e2C197SGS.rv	gctcagcatgccgttgtcggaaccagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with a GS-linker and the sequence adjacent to the fourth extracellular loop of OmpA
d4SGY1e2C197SGS.ds	ccagtggaccaacaactctggtcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttcttttgatcagttcccgtccgattctcaccgtctgtctggttccgacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the SGY1e2C197SGS sequence replacing the fourth extracellular loop of OmpA
d4SGSGY1e2C197S.fw	ccagtggaccaacaactctggtagcggcaggttatgaccgacg	solid phase	desalted	forward primer for flanking the Y1e2C197S sequence with a SGSG-linker and the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e2C197SGSGS.rv	gctcagcatgccgttgtcagagccggaaccagacagacggtgagaa	solid phase	desalted	reverse primer for flanking the Y1e2C197S sequence with a GSGS-linker and the sequence adjacent to the fourth extracellular loop of OmpA
d4SGSGY1e2C197SGSGS.ds	ccagtggaccaacaactctggtagcggcaggttatgaccgacgaaccgttccagaacgttaccctggatgcgtacaaagacaaatacgtttcttttgatcagttcccgtccgattctcaccgtctgtctggttccggctctgacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the SGSGY1e2C197SGSGS sequence replacing the fourth extracellular loop of OmpA

d1Y1e3C294S.fw	ctgggctggtctcagtacttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e3C294S.rv	gcgcccagttggttttcgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with the sequence adjacent to the first extracellular loop of OmpA
d1Y1e3C294S.ds	ctgggctggtctcagtacttcgactggaaccaccagatcatcgcgacctctaaccacaacgaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e3C294S sequence replacing the first extracellular loop of OmpA
d1SGY1e3C294S.fw	ctgggctggtctcagtacagcggcttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with a SG-linker and the sequence adjacent to the first extracellular loop of OmpA
d1Y1e3C294SGS.rv	gcgcccagttggttttcagaaccgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with a GS-linker and the sequence adjacent to the first extracellular loop of OmpA
d1SGY1e3C294SGS.ds	ctgggctggtctcagtacagcggcttcgactggaaccaccagatcatcgcgacctctaaccacaacggtttctgaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the SGY1e3C294SGS sequence replacing the first extracellular loop of OmpA
d1SGSGY1e3C294S.fw	ctgggctggtctcagtactccggtagcggcttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with a SGSG-linker and the sequence adjacent to the first extracellular loop of OmpA
d1Y1e3C294SGSGS.rv	gcgcccagttggttttcgctgccagaaccgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with a GSGS-linker and the sequence adjacent to the first extracellular loop of OmpA
d1SGSGY1e3C294SGSGS.ds	ctgggctggtctcagtactccggtagcggcttcgactggaaccaccagatcatcgcgacctctaaccacaacggtttctggcagcgaaaaccaactgggcgc	PCR	spin column	double-stranded insertion primer for the introduction of the SGSGY1e3C294SGSGS sequence replacing the first extracellular loop of OmpA
d2Y1e3C294S.fw	cgaatggttaggtcgtatgttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e3C294S.rv	cagttgaacgccctgagctttgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with the sequence adjacent to the second extracellular loop of OmpA
d2Y1e3C294S.ds	cgaatggttaggtcgtatgttcgactggaaccaccagatcatcgcgacctctaaccacaacaaagctcagggcgttcaactg	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e3C294S sequence replacing the second extracellular loop of OmpA
d3Y1e3C294S.fw	gtatggcgggccgacacttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e3C294S.rv	gcgaagaccggagaaacgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with the sequence adjacent to the third extracellular loop of OmpA
d3Y1e3C294S.ds	gtatggcgggccgacacttcgactggaaccaccagatcatcgcgacctctaaccacaacgtttctccggtcttcgc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e3C294S sequence replacing the third extracellular loop of OmpA
d4Y1e3C294S.fw	ccagtggaccaacaacttcgactggaaccacc	solid phase	desalted	forward primer for flanking the Y1e3C294S sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e3C294S.rv	gctcagcatgccgttgtcgttgtggttagaggtc	solid phase	desalted	reverse primer for flanking the Y1e3C294S sequence with the sequence adjacent to the fourth extracellular loop of OmpA
d4Y1e3C294S.ds	ccagtggaccaacaacttcgactggaaccaccagatcatcgcgacctctaaccacaacgacaacggcatgctgagc	PCR	spin column	double-stranded insertion primer for the introduction of the Y1e3C294S sequence replacing the fourth extracellular loop of OmpA